

Changing Semantic Role Representations with Holographic Memory¹

Simon D. Levy

Computer Science Department
Washington and Lee University
Lexington Virginia 24450, USA

Abstract

Semantic roles describe “who did what to whom” and as such are central to many subfields of AI and cognitive science. Each subfield or application tends to use its own “flavor” of roles. For analogy processing, logical deduction, and related tasks, roles are usually specific to each predicate: for *loves* there is a LOVER and a BELOVED, for *eats* an EATER and an EATEN, etc. Language modeling, on the other hand, requires more general roles like AGENT and PATIENT in order to relate form to meaning in a parsimonious way. Commitment to a particular type of role makes it difficult to model processes of change, for example the change from specific to general roles that seems to take place in language learning. The use of semantic features helps solve this problem, but still limits the nature and number of changes that can take place. This paper presents a new model of semantic role change that addresses this problem. The model uses an existing technique, Holographic Reduced Representation (HRR) for representing roles and their fillers. Starting with specific roles, the model learns to generalize roles through exposure to language data. The learning mechanism is simple and efficient, and its scaling properties are well-understood. The model is able to learn and exploit new representations without losing the information from existing ones. We present experimental data illustrating these principles, and conclude with by discussing some implications of the model for the issues of changing representations as a whole.

KEYWORDS: *Representation change, learning, development, semantic roles, distributed representations*

The Centrality of Semantic Roles

In modeling language and thought, perhaps no other concept has been as crucial as that of semantic roles. From practically-motivated work in artificial intelligence (Schank 1972) to philosophical discussions of metaphor (Lakoff & Johnson 1980) to theory-driven approaches in linguistics (Fillmore 1968), the importance of representing “who did what to whom” has never been in dispute.

¹This paper is an extended version of an abstract accepted for presentation at the Workshop on Analogies: Integrating Multiple Cognitive Abilities (AnICA07), in Nashville, TN.

The manner in which such roles are specified depends largely on the phenomena of interest, as well as the implementation architecture. Models built on predicate calculus – including general-purpose tools like the Prolog programming language (Clocksin & Mellish 1994) and systems built to solve more specific tasks like analogy processing (Falkenhainer, Forbus, & Gentner 1990) – rely mainly on positional notation; *e.g.*, `loves(mary, john)`, `sees(bill, fred)`. Combined with variable binding, this approach supports logical rules like

```
jealous(X,Y) :- desires(X,Z), desires(Z,Y).
```

that are specific to a given set of predicates.

Some neurally-inspired models, such as (Hummel & Holyoak 2005) and (Eliasmith & Thagard 2001), also make use of predicate-specific roles. In such models these propositions would be represented as `loves(LOVER=mary, BELOVED=john)`, `sees(SEER=bill, SEEN=fred)`, with each capital-letter role implemented by a different unit or set of units, and the equal sign representing some sort of role/filler binding operation (discussed in more detail below). Independent of the notational convention or implementation, the overall tendency is that cognitive models of analogy, deduction, and related processes use roles specific to predicates and relations.

From a linguistic perspective, however, the story is quite different. In addition to the theoretical interest in characterizing the nature and relationships of thematic roles across the world’s languages (Fillmore 1968), there is a practical advantage to constraining the number of roles posited for a given language: without abstract roles like AGENT, PATIENT, INSTRUMENT, *et al.*, it is difficult to express generalizations relating form to meaning. For example, any rule-based system for generating or understanding English declaratives will need to express the generalization that the default order is AGENT-PREDICATE-PATIENT. With roles specific to each predicate, a different rule has to be stated for each new predicate. It is, however, clear that

language users can exploit roles to build new constructions, even without any specific information about the meaning of a verb. Told that *Mary glorped John*, English speakers can construct related sentences like *John was glorped by Mary*, *Who did Mary glorp?*, etc. It is not surprising, therefore, that AI models of language understanding have relied on general roles like these from the very beginning (Schank 1972).

Evidence for Changing Role Representations

The present work takes its inspiration from the framework of Cognitive Grammar (Langacker 1987), with its emphasis on general cognitive mechanisms and learning from examples. In such a framework, one sort of role conceptualization is seen as developmentally prior to, but crucially related to, the other. Because children can reason about the world well before they acquire adult-level language competence, many cognitive grammarians (notably (Tomasello 1992)) favor a view in which general roles emerge from predicate-specific roles through exposure to language. Although this is the approach taken in the experiments described below, it is important to emphasize that the main issue is relating and integrating the two sorts of roles using a single mechanism. The model makes no commitment to a particular direction of change (specific \rightarrow general or general \rightarrow specific), and can accommodate either.

In English and many other languages, information about thematic roles is conveyed primarily through word order (*dog bites man* vs. *man bites dog*). As detailed in (Tomasello 1992), adults' consistent use of word order provides the language learner with an opportunity to generalize roles across different predicates, once the learner has acquired the concepts associated with particular words, plus the general lexical categories *noun* and *verb*. Repeated examples of correspondences between representational components (*noun/filler*, *verb/predicate*) supports integration of different structures into a single conceptual representation.

For example, a child learning English will be provided with numerous consistent associations between simple declarative sequences of the form

$$\textit{noun}_i \dots \textit{verb}_j \dots \textit{noun}_k \quad (1)$$

and conceptual structures of the form

$$\textit{predicate}_j(\textit{role}_{j,l} = \textit{entity}_i, \textit{role}_{j,m} = \textit{entity}_k) \quad (2)$$

The child eventually induces the generalized form

$$\textit{predicate}_j(\textit{role}_l = \textit{entity}_i, \textit{role}_m = \textit{entity}_k) \quad (3)$$

where *role_l* is AGENT and *role_m* PATIENT. A similar story obtains for “free word order” languages like Russian and Japanese, with noun endings taking the place of word order in conveying thematic roles.

Issues in Representation

Cognitive models are often categorized in terms of the *connectionist* vs. *symbolic* distinction. In addition to being descriptively unhelpful, these terms are also typically conflated with a host of issues that may have nothing to do with the commitments entailed by a particular model.

A more useful distinction among cognitive representations is whether they are *local* or *distributed* (van Gelder 1999). Traditional symbol systems (grammar, predicate calculus) use local representations: a given symbol has no internal content and is located at a particular address in memory. Although well-understood and successful in a number of domains, such representations are biologically unrealistic and suffer from brittleness. The number of possible items to be represented is fixed at some arbitrary hard limit, and a single corrupt memory location or broken pointer can wreck an entire structure. Similar issues arise in feature-based representations like Head-Driven Phrase Structure Grammar (Pollard & Sag 1994), and in connectionist models based on “grandmother cell” representations, like temporal synchrony networks (Shastri & Ajjanagadde 1993).

In a distributed representation, on the other hand, *each entity is represented by a pattern of activity distributed over many computing elements, and each computing element is involved in representing many different entities* (Hinton 1984). In addition to being more obviously similar to the way that the brain seems to represent concepts, such representations have a number of appealing properties for cognitive modeling (McClelland, Rumelhart, & Hinton 1986): they are robust to noise, provide realistically “soft” limits on the number of items that can be represented at a given time, and support distance metrics. These properties enable fast associative memory and efficient comparison of entire structures without breaking down the structures into their component parts.

Distributed Representation of Role/Filler Bindings

The best known distributed representations approach is probably the Latent Semantic Analysis (LSA) model of word meaning (Landauer & Dumais 1997), which has proven successful in many application domains. LSA is not however a model of role/filler binding, for which a number of distributed representation approaches have been developed. The approaches that have enjoyed the most success recently can be grouped under the heading *Vector Symbolic Architectures*, or VSAs (Gayler 2003). VSAs trace their origin to the tensor product representations presented in (Smolensky 1990). Tensor-product models represent both fillers and roles as vectors of binary or real-valued numbers. Binding is implemented by taking the tensor (outer) product of a role vector and a filler vector, resulting in a

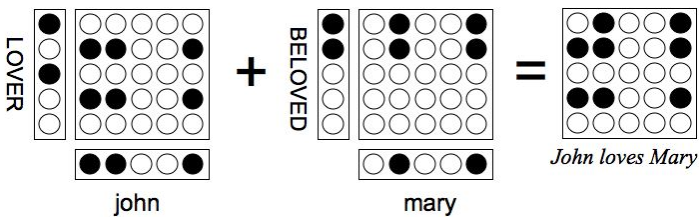


Figure 1: Tensor product representation of *John loves Mary*

mathematical object (matrix) having one more dimension than the filler. Given vectors of sufficient length, each tensor product will be unique. *Bundling* – the combination of role/filler bindings into a complete proposition – can be implemented as element-wise addition (Figure 1). Bundled structures can then be used as roles, opening the door to recursion. To recover a filler (role) from a bundled tensor product representation, the product is simply divided by the role (filler) vector.

Because the dimension of the tensor product increases with each binding operation, this method suffers from the well-known curse of dimensionality (Bellman 1961). As more binding is performed, the size of the representation grows exponentially. The solution is to collapse the $N \times N$ role/filler matrix back into a length- N vector. The most popular approach to doing this is called Holographic Reduced Representations, or HRR (Plate 1991), in which the sum of each diagonal is taken, with wraparound (circular convolution) keeping the length of all diagonals equal (Figure 2). In HRR, roles and fillers are represented as very large ($N > 1000$ elements) vectors of random values drawn from a fixed set or interval. Despite the size of the vectors, the HRR approach is computationally efficient, requiring no costly backpropagation or other iterative algorithm, and can be done using the Fourier Transform, which is $O(N \log N)$. The price paid is that the crucial operations (circular convolution, vector addition) are a form of lossy compression that introduces noise into the representations. The introduction of noise requires that the unbinding process employ a cleanup memory to restore the fillers to their original form. The cleanup memory can be implemented using Hebbian auto-association, like a Hopfield Network (Hopfield 1982) or Brain-State-in-a-Box model (Anderson *et al.* 1977). In such models the original fillers are attractor basins in the networks dynamical state space. These methods can be simulated by using a table that stores the original vectors and returns the one closest to the noisy version.

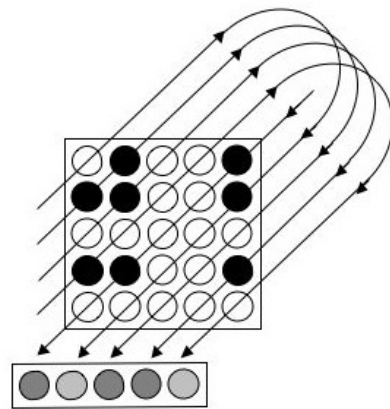


Figure 2: Holographic Reduced Representation of the structure in Figure 1

When everything (predicates, roles, fillers) can be represented with a set of uniformly-sized objects (vectors), all sorts of information about a given structure can be literally added into the representation of that structure, through element-wise addition. For example, the proposition expressed by *Mary loves John* can be represented as

$$\text{JOHN} + \text{MARY} + \text{LOVES} + \text{LOVER} \otimes \text{MARY} + \text{BELOVED} \otimes \text{JOHN} \quad (4)$$

where each capitalized word stands for the vector representation of the corresponding role or filler, $+$ is vector addition, and \otimes is the binding operation. Bundling the representations of JOHN, MARY, and LOVES together with the role/filler bindings supports efficient similarity-based retrieval of surface information (who was involved? what was done?) without the need to perform the unbinding operation. (This is necessary because the binding operation results in a vector that bears no similarity to its role or filler components.) By using bundles as fillers, recursive structures of arbitrary complexity can be built, subject to the “soft limit” phenomenon mentioned above.

HRR has been used successfully in modeling analogy, with results similar to those found in experiments with human subjects (Eliasmith & Thagard 2001). Because the concept of binding is general enough to encode sequencing information (absolute position of a word or its position relative to another), HRR has also been used to model word order in a psychologically realistic way (Jones & Mewhort 2007).

Learning Thematic Roles with Holographic Representations

We conducted three simple experiments to simulate learning of thematic roles from word order. These experiments

tested the ability of an HRR-based model to incorporate thematic roles into a fixed set of predicates and arguments, while maintaining the information from predicate-specific roles useful for other cognitive tasks.

The training set consisted of HRR's for 50 two-place predicates like *LOVES* and *SEES*, and 10 fillers like *JOHN* and *MARY*. A simplifying constraint was that no predicate was allowed to have the same filler in both roles. The arrangement resulted in 4500 ($= 50 \times 10 \times 9$) unique propositions. Predicates, fillers, and roles were represented as 3000-dimensional vectors of Gaussian-distributed random numbers with mean of zero and a standard deviation of $\sqrt{1/3000}$, following the suggestion in (Plate 1991). Propositions were constructed as in example (4) above. To simulate exposure to word-order consistencies as in (1) and (2) above, we used a random vector v_1 to represent the order *noun ... verb* and another vector v_2 to represent *verb ... noun*. The initial propositional HRR structures were rebuilt by replacing the predicate-specific first role with itself plus v_1 , and the second role with itself plus v_2 , and re-binding these summed roles to their fillers.

Experiment 1: Incorporating Thematic Roles

To model the incorporation of thematic roles, we took the vector average of all the integrated agent roles, thereby obliterating predicate-specific information. This average agent role was used as the cue to retrieve a filler from the HRR of each proposition. We did the same for the patient roles. Successful retrieval implies the ability to use a given proposition in various grammatical constructions, including ones to which the model has not been exposed. Retrieval was determined to be successful when the vector cosine of the (noisy) retrieved filler was highest with the original filler used to build the HRR for the proposition (as opposed to any of the other 10 possible fillers). Because we required both fillers to be retrieved successfully, chance performance on this task is 1%. In 10 trials with different sets of random vectors, we achieved a mean success of 99.8%, with a standard deviation of 0.5%. This result shows that the fillers of thematic roles were successfully recovered despite these roles being integrated with the predicate-specific roles.

Experiment 2: Does Representation Change Hurt Performance?

As discussed earlier, many kinds of cognitive processing require representations of propositional content that support querying the specific, concrete role played by a given entity. A representation of *Mary loves John*, for example, must report Mary's role as *LOVER*, and not merely something general like thematic role *AGENT*. To test the retrieval of these predicate-specific roles after the representation change to thematic roles, we used the fillers in each proposition as retrieval cues. As in the experiment above, the results were

highly successful: both of the concrete roles were retrieved correctly 97.5% of the time, with a standard deviation of 5% over the 10 trials (chance $\ll 1\%$). As in the first experiment, the integration of both role components into a single representation did not adversely affect the ability to retrieve crucial information. The model supports both linguistic and non-linguistic processes in a single, coherent representation.

Discussion

Criticism of distributed representations and other neurally-inspired approaches has come from two directions. Some cognitive scientists have claimed that such representations cannot arrive at or exploit systematic, compositional structure of the sort used in traditional cognitive science and AI (Fodor & Pylyshyn 1988). A minimum requirement noted by these critics was that a model that could represent the idea *John loves Mary* should also be able to represent *Mary loves John* (systematicity) and to represent *John*, *Mary*, and *loves* individually in the same way in both (compositionality).

From the preceding discussion it should be clear that Holographic Reduced Representations (and other Vector Symbolic Architectures) address these issues – specifically, by providing operators (binding, bundling) that compose structure in a systematic way, instead of treating on compositionality and systematicity as emergent properties.

A second set of criticisms has come from proponents of traditional AI (Minsky & Papert 1988). Such critics worry about the assignment of credit to the right component of a system in which information is distributed among all representational elements, as well as the ability of such systems to scale up to larger, real-world applications.

Dealing with the credit-assignment problem will likely require a more sophisticated notion of “component” than the one offered in the traditional AI literature. Figure 3 shows a principal components analysis (Shlens 2005) of role vectors from the first two experiments, before and after thematic role incorporation. (The axes correspond to the two dimensions with the highest eigenvalues.) As the middle image in the figure suggests, there is indeed a strong internal structure to the HRR representations, which reveals how they support the behavior found in these experiments. Qualitatively, there are two broad role categories. A closer look at this structure (rightmost image) reveals a level of detail beyond the high-level agent/patient distinction visible in the middle image – quantitatively, there are still as many role categories as there were originally. Although this data was admittedly constructed to have exactly these properties, the general point remains valid: to move forward, AI will need new models of representation that support structure at many levels of detail, and that support representation change without re-engineering.

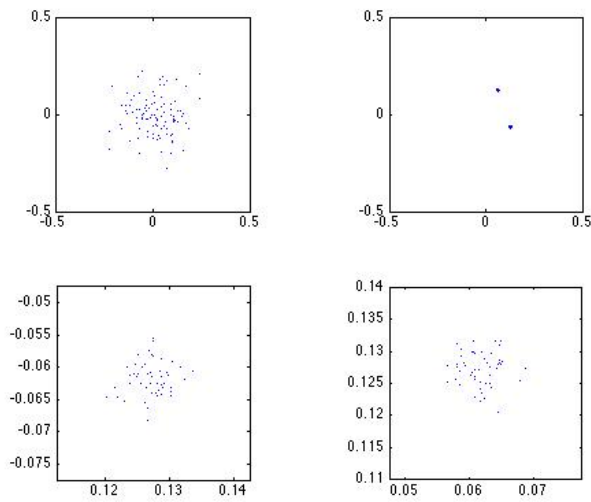


Figure 3: Principal Components Analysis of role vectors. Top left: before addition of theta roles. Top right: after addition of theta roles. Bottom: detail from top right.

A precise treatment of the scalability issue is well beyond the scope of the present work. The appendices in (Plate 2003) provide an exhaustive analysis of the capacity of HRR and convolution-based memories in general. The two main results of this analysis are (1) a linear increase in capacity as the length of the vectors increases and (2) an exponential increase in the required length as the average similarity of the vectors increases. Together, these results suggest that vector-based architectures will benefit as much as traditional architectures from Moore’s-Law increases in memory and processor speed, while offering psychologically more realistic behavior (confounding of similar items, short-term memory effects, etc.).

Conclusions and Future Work

This paper has presented some preliminary results in using a novel computational medium, holographic / convolution memory, to model a important of representational change: the change from specific to general semantic roles that seems to take place in language learning. Crucially, the results show that this medium can accommodate the new representation without hurting the existing one. As compared with traditional discrete (localist) symbol systems like grammars, holographic memory offers a more supple, psychologically plausible basis for representing this kind of change. The ability to seamlessly integrate linguistic information into existing, pre-linguistic representations opens the door to exploring longstanding hypotheses on the possibility that language can influence thought (Whorf 1939).

There are several pieces are missing from our current story. One question not addressed in the experiments is the issue of generalization to unseen examples. Consider, for example, the situation of a first-language learner who has pre-linguistic mental representations for a number of

action predicates and their roles, and who has heard linguistic exemplars for similar predicates. There is ample evidence that learners generalize, and even over-generalize, thematic roles in these situations: Bowerman (1982) provides examples like *Don’t you fall me down*. Because the notion “similar” has a precise mathematical interpretation in a vector-space representation, it should be possible to use vector-space learning methods, like the Self-Organizing Map (Kohonen 2001), to model this sort of generalization.

A second, related question concerns the genesis of the predicates and roles themselves. We have simply assumed the existence of a “lexicon” of predicates, each with a set of roles, without worrying about how such elements might emerge from the pre-linguistic experience of an individual – *i.e.*, from perception and action in the world. A longstanding tradition in philosophy, stretching from Plato through Descartes to Chomsky (1966), bypasses this question by assuming the existence of “innate ideas” that each human being is born with. Recently, some cognitive scientists (*e.g.* Steedman 2002) have begun to explore the more interesting possibility of relating predicate/role representations to the theory of affordances in animal behavior (Gibson 1977). We might speculate on the possibility of a system that would take as input a visual representation of an event and a visual representation of one of the participants, and would return a representation of the role played by the participant.

Also missing in our story is an explicit mechanism for transducing role representations to/from word sequences. Modeling these mechanisms in a neurally realistic manner is an active research topic (Dominey, Hoen, & Inui 2006). The work presented here represents a first step toward integrating such “surface-level” information with a sufficiently rich semantic representation in a neurally plausible way.

Acknowledgments

The author thanks Ross Gayler, Harry Howard, Pentti Kanerva, and Tony Plate for inspiration and useful suggestions on an earlier version of this paper.

References

- Anderson, J.; Silverstein, J.; Ritz, S.; and Jones, R. 1977. Distinctive feathers categorical perception and probability learning: some applications of a neural model. *Psychological Review* 84(5):413–451.
- Bellman, R. 1961. *Adaptive Control Processes*. Princeton University Press.
- Bowerman, M. 1982. Evaluating competing linguistic models with language acquisition data: Implications of developmental errors with causative verbs. *Quaderni di semantica* 3.

- Chomsky, N. 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper and Row.
- Clocksink, W., and Mellish, C. 1994. *Programming in Prolog*. Berlin: Springer Verlag.
- Dominey, P.; Hoen, M.; and Inui, T. 2006. A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience* 18:2088–2107.
- Eliasmith, C., and Thagard, P. 2001. Integrating structure and meaning: a distributed model of analogical mapping. *Cognitive Science* 25(2):245–286.
- Falkenhainer, B.; Forbus, K. D.; and Gentner, D. 1990. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence* 41:1–63.
- Fillmore, C. 1968. The case for case. In Bach, E., and Harms, R., eds., *Universals in Linguistic Theory*. Holt, Rinehart. 1–90.
- Fodor, J., and Pylyshyn, Z. 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition* 28:3–71.
- Gayler, R. 2003. Vector symbolic architectures answer jackendoffs challenges for cognitive neuroscience. In Slezak, P., ed., *ICCS/ASCS International Conference on Cognitive Science*, CogPrints, 133–138. Sydney, Australia: University of New South Wales.
- Gibson, J. J. 1977. The theory of affordances. In Shaw, R., and Bransford, J., eds., *Perceiving, acting, and knowing: Toward an ecological psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates. 67–82.
- Hinton, G. 1984. Distributed representations. Technical Report CMU-CS-84-157, Computer Science Department, Carnegie Mellon University.
- Hummel, J. E., and Holyoak, K. J. 2005. Relational reasoning in a neurally-plausible cognitive architecture: An overview of the lisa project. *Current Directions in Psychological Science* 14(3):153–157.
- Jones, M. N., and Mewhort, D. J. K. 2007. Representing word meaning and order information in a composite holographic lexicon. *Psychological Review* 114.
- Kohonen, T. 2001. *Self-Organizing Maps*. Secaucus, NJ: Springer-Verlag, 3 edition.
- Lakoff, G., and Johnson, M. 1980. *Metaphors We Live By*. University of Chicago Press.
- Landauer, T. K., and Dumais, S. T. 1997. A solution to plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 104:211–240.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar*. Stanford University Press.
- McClelland, J.; Rumelhart, D.; and Hinton, G. 1986. The appeal of parallel distributed processing. In Rumelhart, D., and McClelland, J., eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. MIT Press.
- Minsky, M. L., and Papert, S. A. 1988. *Perceptrons*. MIT Press.
- Plate, T. 1991. Holographic reduced representations. Technical Report CRG-TR-91-1, Department of Computer Science, University of Toronto.
- Plate, T. A. 2003. *Holographic Reduced Representation: Distributed Representation for Cognitive Science*. CSLI Publications.
- Pollard, C. J., and Sag, I. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Schank, R. C. 1972. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology* 3(4):pages 532–631.
- Shastri, L., and Ajjanagadde, V. 1993. From simple associations to systematic reasoning. *Behavioral and Brain Sciences* 16(3):417–494.
- Shlens, J. 2005. A tutorial on principal component analysis. Technical report, Systems Neurobiology Laboratory, Salk Institute for Biological Studies, <http://www.snl.salk.edu/shlens/pub/notes/pca.pdf>.
- Smolensky, P. 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46:159–216.
- Steedman, M. 2002. Plans, affordances, and combinatory grammar. *Journal Linguistics and Philosophy* 25(5-6).
- Tomasello, M. 1992. *First Verbs: A Case Study of Early Grammatical Development*. Cambridge University Press.
- van Gelder, T. 1999. Distributed versus local representation. In *The MIT Encyclopedia of Cognitive Sciences*. Cambridge, Massachusetts: MIT Press. 236–238.
- Whorf, B. L. 1939. The relation of habitual thought and behavior to language. In Duranti, J. A., ed., *Linguistic Anthropology: A Reader*. Oxford, UK: Blackwell Publishers.