Incremental Modeling of Language Understanding Using Speech Act Frames

Wende Frost (wende.frost@asu.edu)
Arizona State University
School of Computing and Informatics, 699 S. Mill Avenue
Tempe, AZ 85281 USA

Magdalena Bugajska (magda.bugajska@nrl.navy.mil)
J. Gregory Trafton (greg.trafton@nrl.navy.mil)
Navy Center for Applied Research in Artificial Intelligence
Naval Research Laboratory, 4555 Overlook Avenue SW
Washington, DC 20375 USA

Abstract

Using the cognitive architecture ACT-R/E, we designed a framework for implementing cognitively plausible spoken language understanding on an embodied agent using incremental frame representations for multiple levels of linguistic knowledge. Emphasis is placed on semantics, pragmatics, and speaker intent.

Keywords: cognitive modeling; language processing; pragmatics; semantics

Introduction

One of the greatest challenges in building and interacting with embodied agents is integrating cognitive plausibility without sacrificing usability. This challenge is very evident when looking at natural language understanding in spoken language environments. In real-life scenarios, agents need to respond to commands, gather information, and answer queries quickly even when faced with unexpected input from human users. Unexpected input can occur at many different linguistic levels, whether due to the agent's speech recognition software failing to recognize a word, the need to process and understand an irregular syntactic utterance, or by verbal interruptions in the middle of a task. While failure to promptly cope with all kinds of unexpected input leads to an agent being less useful in the field, it also exposes a lack of cognitive plausibility in the framework of the model. Humans do not reach long-lasting impasses when faced with any of these relatively simple situations (Gibson, 1991). In addition to performing such tasks, the agent should also be able to hear, process, and remember utterances the speaker directed at other agents in the environment without mistaking them for commands or queries the speaker expected the agent to achieve. A cognitively plausible agent would then be able to use these utterances directed toward others, especially relatively recent ones, and incorporate them with world knowledge for use in future goals.

We have implemented a framework within the ACT-R/E 6 cognitive architecture (Anderson et al., 2004; Anderson & Lebiere, 1998) which aims to fulfill these requirements at both functional and plausible levels. The framework's focus is to obtain a correct interpretation of the speaker's intentions (i.e. what the speaker wants) based upon the current state of the world and what existing past world knowledge a model has in memory, rather than a syntactically exact parse of the utterance (i.e. what the speaker said) divorced from an outside environment. Processing in our framework is done at all levels for each word as it comes in. It retrieves, creates, and edits frames of knowledge from the phonological to the pragmatic at each step, enabling an agent to have a constantly developing picture of the utterance.

Our framework differs from other natural language work in the ACT-R family of architectures by focusing on processing spoken natural language in real-time and on emphasizing pragmatic and semantic roles in the utterance. It was also created to be easily expandable in other useful embodied directions, including processing gestural information as part of an utterance structure. In keeping with the fundamental notion that language processing is another aspect of human cognition subject to the same representations and processes as other cognitive activities, our framework does not implement a dedicated "language module." (Croft, 2004) Instead, language processing is done across existing modules. This non-dedicated module approach differentiates our framework from much other ACT-R work on language, such as Ball (2007) and Emond (2006). As with other work in ACT-R that does not have a dedicated language module (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006), we have included an additional buffer to store language information. Unlike this work, however, we do not have any parallel lexical access mechanisms. Our divergence from a modular approach is also similar to the NL-SOAR language comprehension theory implemented in the SOAR cognitive architecture, but NL-SOAR focuses on explaining a large number of sentence-level syntactic phenomena (Lewis, 1993), whereas we place more emphasis on semantics and pragmatics.

To achieve a cognitively plausible framework to model natural language understanding, we used the ACT-R/E cognitive architecture with the default ACT-R parameters set. ACT-R 6 is a production system architecture composed of two kinds of knowledge: declarative and procedural. Declarative knowledge, also known as factual knowledge, is stored in long term declarative memory as "chunks." These chunks, as well as chunks based upon perceptual information gained from sensors, are retrieved into central cognition by way of "buffers." Procedural knowledge is a set of condition-action rules, from which one is chosen to fire after a conflict-resolution process based on the expected gain to alter the state of the buffers. Different chunks have different levels of activation affecting time taken for retrieval. Standard ACT-R interfaces with the outside world through visual, aural, motor, and vocal modules. The architecture supports other faculties through intentional, imaginal, temporal, and declarative modules.

ACT-R/E is a modified version of ACT-R that allows the architecture to perceive the physical world by attaching robotic sensors and effectors to it. It includes a new module (spatial) and modifications to the visual, aural, and motor modules to work with our robot and to use real-world sensor modalities. The rest of the architecture was not modified.

Language Representation

From the time a model hears a sound to the time a model is acting on a fully processed utterance, we have identified four major representation types; content of sound, meaning, phrase, and speech act, based on Clark (1996). To represent these types, we have used a form of frame representations (Langacker, 1999) which integrates well with the format of chunks in ACT-R/E. Each frame consists of an identifier and a type, followed by a list of slots that will later be matched against by the productions.

Simply using frames or schemas to represent linguistic expressions appears similar to the family of Construction Grammar theories (Fillmore & Kay, 1993; Goldberg, 1995), but our framework differs from Construction Grammar in fundamental ways, including the recognizance of synonymy.

Due to the structure of our representations, the same representation will work for languages besides English, including most analytic, SVO, head-first languages. Our representation is currently limited to languages with structural differentiation between grammatical moods.

Content of Sound

The content of sound representation is an existing ACT-R/E construct. Sound events are processed in the aural module, with the sound location in the aural-location buffer and the sound content in the aural buffer. The sound content in the aural buffer on the first word of the sentence "Go to the corner office by the lobby," is represented in Figure 1.

```
[word0-0     isa          sound
             kind         word
             content      go
             event        audio-event1]
```

Figure 1: Word Sound Frame

Meaning Frames

The frame of meaning representation is the first type of chunk our framework tries to retrieve when it knows it has a sound of content kind "word." This is the difference between a model likely understanding the meaning of a word and a model dismissing a word not in its vocabulary. When it does not find at least one meaning chunk for a sound in its declarative memory, the sound is dismissed as a completely unknown word with no need for further processing. The framework does not try to fit unknown words into higher-level representations, which makes the recovery time very fast for a lexical error or filler word. The activation level of the 3000 most commonly used words in the English language has been set very high. All words and word senses are hand-entered, as they will be until the definitional frames are regularized.

The most basic meaning frame, as shown in Figure 2, consists of a unique identifier for the word sense, the general *identifier* for the lexeme, the *part of speech*[1] of the word, and whether or not this word is a *catalyst* for changing to a new phrase frame. Catalysts include verbs, some types of nouns (e.g. vocative), prepositions, conjunctions, and complementizers.

Currently, the catalyst slot can be filled by three values: "yes," "no," and "nil." A catalyst of "yes" indicates a change to a new phrase frame. A catalyst of "no" indicates that a word is a content word and should be integrated with the rest of the utterance. A catalyst of "nil" indicates that the word should either be disregarded immediately, or after a minor change is made to a value in an existing frame.

```
[to-1        isa          meaning
             identifier   to
             pos          locative-prep
             catalyst     yes]
```

Figure 2: Basic Meaning Frame

In addition to the basic items listed above, the meaning frame also contains the necessary agreement information about verbs and nouns, such as tense, plurality, person, etc. Definitional information about function words will be stored as separate types of chunks.

Phrase Frames

Once a model has a meaning frame for a word, it next fills in a phrase frame where appropriate with the meaning. The phrase frames are composed of a loosely linked set of information from the meaning frames compiled into a higher-level semantic structure. The same phrase frame is used for incoming meaning until a word functioning as a catalyst is heard. This means that each phrase has several optional slots in the frame to accommodate the different configurations of phrases in natural language and only two basic slots. The basic slot is the *phrasetype* slot, which is filled with the type of phrase being created. In addition, the phrase frame usually contains the head of phrase, or the

[1] Verb and noun types are stratified based upon their WordNet classifications.

word which triggered the phrase change, any modifiers to the phrase itself, what thematic role the modifier plays, any thematic words in the phrase, as well which thematic role they play, and any modifiers to the thematic words in the phrase.

[locative	isa	phrase
	phrasetype	locative-prep
	head	to
	modifier	by
	modifier-role	locative-prep
	thematic	office
	thematic-role	destination
	thematic-mod1	corner]

Figure 3: Basic Phrase Frame

The phrase frame listed in Figure 3 shows the phrase headed by "to" in the utterance "Go to the corner office by the lobby now." Since "to" is a catalyst, a new frame was created containing it as the head. As previously noted, non-content words that have little bearing on the eventual intention of the utterance (such as "the" in this example) are not stored in a slot. As in Altmann (1999), thematic roles for the slots are filled partially based on context. The first and only thematic role related to locative prepositions is destination. Since "office" does not fall in the role of catalyst or in the role of non-content word, it is seen to play a thematic role in the sentence, filling in the expected role of "destination." As soon as the word "by" is uttered with its catalyst value of "yes," a new phrase frame is created to hold the information "by the lobby now."

Phrase frames hold much of the semantic content of the utterance, as well as the syntactic linkages. They do not contain any immediate method of unifying the phrases created. The goal of the framework is not to immediately have perfect recall of utterance parses and integrations at all levels, but rather to derive the intention of the utterance. This goal is plausible based on Langacker (1999).

Speech Act Frames

Speech act frames, roughly based on the speech acts of Austin (1962) and Searle (1969), are composed over the course of the utterance with a recognized cue creating a barrier between utterances[2] and ending the speech act. For every new addition or alteration to a phrase, there are productions to update the current speech act frame. As many speech acts are composed of multiple utterances or sentences, there will have to be another level of composition added in the future to distinguish between the intention of single utterances and the intention of the speech act as a whole.

The speech act frame has three basic slots: *type*, *actual state*, and *desired state*. *Type* is filled with the appropriate modality for that speech act: "indicative," "imperative," or "declarative." Other moods, such as subjunctive or jussive, are not recognized at this time due to difficulty in recognizing their structural components. The currently recognized moods form the set of recognized structural pragmatic markers (Fraser, 1990). Other discourse markers, such as phrasal patterns (Pitler, et al., 2008; Saito, Yamamoto & Sekine, 2006), will be integrated later, adding to the possible values in the *type* slot.

Actual state is based on a model's perceptual and declarative knowledge of the state of the world. *Desired state* stores any speaker intent as to the desired state of the world, which is gleaned from their utterances. Comparing the two states is useful when following commands or checking the mood of a statement.

The other content of the speech act frame includes *actiontype*, *who*, *how*, *when*, *where*, *what*, and any more specific values, such as *where-exactly*, that are needed to understand an intention. They were named to be as easy as possible for the human user to interpret, as this is the frame level used by a human to check for understanding of intent. The *actiontype* is the high level action the agent needs to take, such as implementing a verb of motion. The specificity of verb types can be tailored to the capabilities of the robot. Limiting the *actiontype* field to types of actions (e.g. verb of motion) rather than specific actions (e.g. walk) in every case allows for easier analysis of unknown verbs. The verb itself is stored in the *how* slot in case there is an action or production based on a specific verb.

The *who* slot contains who the utterance was directed toward. By default, all utterances are assumed to be directed at our robot unless specifically stated otherwise. Similarly by default, all *when* slots are assumed to be "now" unless another modifier is given, such as "in two hours" or "after." The *where* slot is filled with destination or source information gleaned from the phrase frames, and the *what* slots refer to patients, instruments, recipients and other thematic roles not already covered. Unlike phrase frames, speech act frames can contain multiple lexemes in the same slot, such as "corner office."

An example of the speech act frame from the command for the robot to walk to the corner office by the lobby is given in Figure 4.

[general	isa	speech-act
	type	imperative
	actiontype	verb-motion-intran
	who	robot
	where	corner office
	where-exactly	by lobby
	when	now
	actual state	n
	desired state	y]

Figure 4: Speech Act Frame

[2] This cue is currently the word "now." With fine-tuning of speech recognition software, we hope to instead use prosodic analysis tools to correctly identify the end of the utterance based on length of pauses correlated with surrounding pitch.

Utterance Processing

The primary goal of our language understanding framework is to have an embodied agent operate in real-time in a cognitively plausible fashion, integrating information from the perceptual modules as well as declarative memory to form a picture of the speaker's intentions. Cognitive plausibility requires both representations and processes to be plausible. Plausibility in the representations is gained through the use of linguistic frames (Langacker, 1999), while processing gains plausibility by matching human perceptual and cognitive data over a series of processing steps.

Audio Processing

The first step of processing spoken language is recognizing incoming sounds. In the framework, a model has productions to continually monitor the environment for sound, even as it's processing or achieving another goal. This is done by attending to the audio events appearing in the aural location buffer. Once an audio event has been attended to, its content is placed in the aural buffer.

We have tested the audio processing of the framework in two different ways: by receiving aural input by means of the commercial speech-recognition engine ViaVoice and by simulating the arrival of audio events in the aural location buffer through text input. The latter method is useful for precise testing of input speed. Though there is evidence that humans interacting with what they know to be non-human systems speak more slowly than usual (Lewis, 1999), when we used the manual input method, we set the event arrivals at the speed of humans interacting with other humans, which varies from 180 to 250 words per minute (WPM) (Picard, 1997). The event arrival was set accordingly at one word per 30 ms. In addition, the default sound decay time value is 3.0 sec. This means that if an audio event is not attended to within that time from its onset, it will become unavailable to a model.

Serial Processing

Since work has been done showing that humans comprehend utterances at all linguistic levels nearly simultaneously (Hagoort, 2008), our framework has the ability to process the different frames of an utterance extremely quickly, with all levels of processing done for each word.

This processing speed is reached by means of the aforementioned addition of the language buffer with a minimum one retrieval and maximum three retrievals per each recognized word heard.

In the most rapid case, one retrieval is necessary to retrieve the meaning of the heard word. The word is either a catalyst of "nil," it is a thematic role to both a phrase and speech act currently being processed, or it is a modifier to a thematic role in an existing phrase. In these cases, the word and its role are written to the existing chunks in the appropriate buffers and sent to declarative memory.

In the intermediate case, after the initial retrieval gains the meaning of the word, another retrieval is necessary for a new phrase type chunk. This case is signaled by the catalyst slot in the meaning retrieved being filled with the value "yes." While this case takes slightly more time, it is theoretically plausible that changing to a new phrase type requires more cognitive effort than filling in slots in an existing phrase type. The speech act frame in this instance is still the same.

In the most extreme serial case, a new utterance is being processed. This requires retrievals not only of the word meaning and phrase type, but of the speech act itself. In this case, however, the simultaneity component is not as relevant since Hagoort's data did not assume any semantic, syntactic, or pragmatic information was present before speech began.

Cognitive Response Time

Speech understanding is normally done at a rate of 150-160 WPM (Williams, 1998) or an average of one word per 40 ms. According to Card, Moran, and Newell (1983), this rate corresponds with the cognitive cycle time for processing information about each word: from 25-170 ms. The utterance understanding time in our framework, 84 ms, was measured between the time the last word was uttered and the time the phrase was fully understood. This fits well within the range given by Card et al. The average WPM rate, 41, was found by dividing the time the utterance was fully understood by the number of words in the phrase. This is very close to the rate found by Williams. The default ACT-R/E retrieval and conflict-resolution values were used. These initial constraints, along with the serial retrieval mechanism of ACT-R/E, contributed to the cognitively plausible cycle times gained.

Data Retrieval

After the word sound has been attended to, it is checked against all word senses in memory to see if a meaning can be retrieved. If a meaning is located, it is retrieved and a speech act is created. Once the speech act is created, a phrase frame is retrieved; the type depends upon the meaning. If the meaning cannot retrieve phrase frames (i.e. it is not marked as a catalyst), it is either placed into the speech-act or into declarative memory. Most words in this situation will be sent to declarative memory, as the speech-act role will not be obvious at the onset.

Once a phrase frame has been retrieved, the slots are filled based upon incoming meanings until a word of a separate or embedded phrase type comes into the system. Though no productions are firing in parallel, there is conflict resolution between the productions to fill slots in the phrase frame and the productions to fill slots in the speech act frame, leading to interleaving of the completion of the two frames as words are heard.

Interruption, Resumption, and Disregard

Once an utterance in the imperative mood is complete, a model checks to see if there are any appropriate productions which support the achievement of this command. If the utterance was unfinished (e.g. important information such as the destination in a movement command was missing), the model will wait for another utterance.

If there is sufficient information encoded to complete the command and a production has been fired to start achieving the command, the slot in the speech act frame for *actual state* will change from "n" to "in process." If the command is interrupted by another command directed to the robot, it will follow the new command instead, and place the interrupted speech act in the declarative memory. If the robot is later told to "never mind," "keep going," or "continue," it will retrieve the interrupted speech act and set about achieving it again. If the robot has been interrupted multiple times, it will only be able to go back to the interrupted command with the highest activation, unless specific information about which interrupted command is desired is given.

Since a speaker may be directing commands or dispensing facts to multiple agents at the same time, the framework only regards commands that have been addressed to it as goals to fulfill. The commands and statements directed at other agents or humans are organized as speech act frames, then put into declarative memory for future use.

Discussion

Our framework operates incrementally in ACT-R/E on simple[3] commands and declarative statements, with representations, perceptual cycles, and cognitive cycles that are cognitively plausible. The framework is reasonably error tolerant at both a lexical and syntactic level, with more attention given to the intention of the speaker than to the preciseness of the input. Some of this flexibility is gained by not relying on syntax for more than clues about the utterance and intention. As spoken word input is often syntactically flawed while remaining semantically coherent, we felt this was a reasonable approach. A slightly higher level of syntactic productions will be added in the future. The focus of the framework is on understanding the intent of the utterance and creating the speech act frame according to the pragmatic information gained by the modal structure.

Since the framework operates in a cognitively plausible cycle time, it is able to analyze and act upon speech acts as they are given in real time, as humans do. There is no backlog of words that "decay" or are "forgotten" before a model analyzes them, due to the high activation of common words, so it can continue virtually indefinitely. The only situation that would result in a model lagging significantly behind the speaker would be that in which more than 50% of the lexical items are unknown or extremely rare.

The framework is able to divert attention from achieving one goal when it is interrupted by another goal. In addition, it can retrieve these interrupted goals when directed to do so. The robot does not act upon any commands other than those the speaker intended, which provides functionality for directing an agent as part of a team. The robot does not start achieving goals given in the commands until the utterance is complete. Future work on priming will give the robot the ability to begin achieving goals directed toward it even before the speaker has finished the utterance.

While not yet as large-scale as language understanding systems such as Ball (2007) and Lewis (1993, 2005, 2006) have created, we feel that by placing minimal reliance on syntax and focusing on semantics and pragmatics, our language framework has the potential to become a worthwhile addition to the field of natural language understanding.

Future Work

Future work on this framework will proceed along two complementary avenues. Both avenues will more fully take advantage of the embodied aspect of the framework. The short term work will concentrate on integrating more technical capabilities with the existing framework and on gathering data for other languages and situations (such as gestural recognition) currently within its capacity. An example of expanded technical capabilities would be incorporating speech-recognition software that contains the prosodic analysis tools for pause and pitch mentioned earlier. Integration with a lexical database, such as WordNet, is also paramount.

Currently, data has been gathered in the framework using simple commands, declaratives, interruptions, and resumptions in English using a variety of verbs, prepositions, and location phrases. To ensure robustness and continued plausibility of the run times, further studies should be run on more complicated or multi-sentential utterances, queries, and non-English domains. None of these examples should require major changes to the existing framework.

Long term work to make a larger scale framework will focus on three major areas: priming for upcoming words and parts of speech, stable left branching, and definitional frame regularization.

The existing framework implementation does not have any priming. This, on top of not being cognitively plausible (McNamara, 2005), hampers it in regard to unknown words. With the addition of priming for upcoming parts of speech or thematic roles, the role of an unknown word in an utterance may be inferred, even if the exact meaning of the word itself remains unknown. This will let the robot query the speaker when there is a barrier to understanding an important part of an utterance, yet continue to discard unknown words that play no key role in a speech act. Priming will also aid in the framework's currently weak word sense resolution.

Since the framework was created using simple English head-first grammar as a template, all productions are currently geared toward right branching phrases. There are only a few instances of left branching permitted through the

[3] "Simple," in this case, means no more than two prepositional phrase embeddings, minimal left branching, and no center embedding or compound utterances.

current productions. While the representations will handle left branching with very few modifications, the productions will need to undergo significant changes. Once left branching is integrated more thoroughly into the framework, it will also be able to process utterances in SOV analytic languages.

Definitional frame regularization will involve defining the different senses of the words in such a way that they can be retrieved by major features or roles held. This will be a fairly substantial undertaking, as even once a regularization is decided upon, which is no small task, there is no guarantee that data from an existing database can be easily altered to fit the chosen chunk format.

In addition to modifying the framework, we will also show that the model executed on the framework has plausible reaction times by dataset matching.

Supplementing these additions and expansions to the framework within ACT-R/E, work will begin towards integrating the underlying principles of the framework with other cognitive architectures, such as SOAR and Icarus (Langley & Choi, 2006). This integration will lend credence to the robustness of the framework while making the framework more accessible to users of different cognitive architectures.

Acknowledgments

This work was supported by the Office of Naval Research under funding document N0001409WX20173 to JGT. The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U. S. Navy.

References

Altmann, G. T. (1999). Thematic role assignment in context. *Journal of Memory and Language*, *41(1)*, 124-145.

Anderson, J.R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111(4)*, 1036-1060.

Anderson, J. R. & Lebiere, C. (1998). *The Atomic Components of Thought*. Lawrence Erlbaum Associates.

Austin, J. L. (1962). *How To Do Things With Words.* Cambridge: Harvard University Press.

Ball, J. (2007). Construction-driven language processing. In *Proceedings of the Second European Cognitive Science Conference*, Delphi, Greece.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Academic Press.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Croft, W. & Cruse, D. A. (2004). *Cognitive Linguistics.* Cambridge University Press.

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* (pp. 359-360). Trieste, Italy.

Fillmore, C. & Kay, P. (1993). *Construction grammar coursebook.* Berkeley, CA: Copy Central.

Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics, 14,* 383-398.

Gibson, E. A. F. (1991). *A computational theory of human linguistic processing: memory limitations and processing breakdown.* PhD thesis, Carnegie Mellon. Available as Center for Machine Translation technical report CMU-CMT-91-125.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Cognitive Theory of Language and Culture. Chicago University Press.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363,* 1055-1069.

Langacker, R. W. (1999). *Foundations of Cognitive Grammar: Volume II: Descriptive Application*. Stanford University Press.

Langley, P. & Choi, D. (2006). A unified cognitive architecture for physical agents. In *Proceedings of the twenty-first national conference on artificial intelligence*. Boston: AAAI Press.

Lewis, J. R. (1999). Effect of error correction strategy on speech dictation throughput. In *Proceedings of the Forty-Third Annual Meeting of the Human Factors and Ergonomics Society* (pp. 457-461).

Lewis, R. L. (1993). An architecturally-based theory of sentence comprehension. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.*

Lewis, R. L. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375-419.

Lewis, R. L., Vasishth, S. and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, *10*, 447-454.

McNamara, T. P. (2005). *Semantic priming*. Psychology Press.

Picard, R. W. (1997). *Affective computing.* Cambridge, MA: MIT Press.

Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., & Joshi, A. (2008). Easily identifiable discourse relations. To appear in *Proceedings of COLING 2008*. Manchester, UK.

Saito, M., Yamamoto, K., & Sekine, S. (2006). Using phrasal patterns to identify discourse relations. In *Proceedings of the Human Language Technology Conference of the NAACL.* (pp. 133-136). New York, NY: Association for Computational Linguistics.

Searle, J. (1969). *Speech Acts.* Cambridge University Press.

Williams, J. R. (1998). Guidelines for the use of multimedia in instruction. In *Proceedings of the Human Factors and Ergonomics Society Forty-Second Annual Meeting* (1447-1451).