

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The online processing of modal verbs: Parallel activation of competing mental models

Permalink

<https://escholarship.org/uc/item/15s724hw>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN

1069-7977

Authors

Huette, Stephanie

Matlock, Teenie

Spivey, Michael

Publication Date

2010

Peer reviewed

The online processing of modal verbs: Parallel activation of competing mental models

Stephanie Huetten (shuette@ucmerced.edu),
Teenie Matlock (tmatlock@ucmerced.edu),
and Michael J. Spivey (spivey@ucmerced.edu)

School of Social Sciences, Humanities & Arts, University of California, Merced, Merced, CA, 95344, USA

Abstract

A simple audio-visual two-alternative forced-choice task was conducted to examine processing differences between the modal verbs *should* and *must*. Unambiguous propositions were either agreed with or disagreed with, and participants' eye movements were monitored as they heard and read the sentence. Reaction times reveal no differences in processing. However, closer time course analyses revealed a divergence in fixations to the target for *should*. These results suggest two mental models are simultaneously activated, entailing both agreement and disagreement with the statement in question.

Keywords: Language understanding, semantics, modal auxiliaries, eyetracking

Introduction

Modal verbs are a central part of modality, the part of language that broadly concerns the factual status of affairs (Frawley, 1992). These verbs are often used to express how the world should be (*deontic* modality) or how the world must be because of some inference (*epistemic* modality) (Sweetser, 1990; Nuyts, 2006). The primary semantic function of modals is to deintensify the meaning of what is stated, by indicating some degree of uncertainty (Close, 1975). Humans are uniquely equipped to think and communicate states of affairs that are ideal, that have never been witnessed first-hand, or that can only exist if some past action were different. For instance, the epistemic use of the modal *must*, as in "You must be bored" indicates an inference (e.g., somebody is nodding off or appears inattentive), while its deontic use, as in "You must pay your taxes" indicates an obligation to perform a series of actions that are ideal according to customs and laws. This initial line of research focuses on the perception of deontic modality. Spoken language is processed incrementally (Marslen-Wilson, 1975), and thus the parallel processing and certainty differences of *should* and *must* may be accessible when temporally sensitive metrics are used.

Modals are of interest to developmental psychologists who investigate reasoning. In one study, children predicted the outcome of a scenario in which a character used a phrase such as "You can't go outside" directed toward a fictitious character (Byrnes & Duff, 1989). Statements including *might*, *have to*, *can*, and *can't*, were used. The results revealed that children ages 3-5 are able to combine negation (*can't*) and modal auxiliary intensity to infer future actions in a story, specifically performing more accurately with *have to*, and less accurately with *might*. Children often responded with the opposite of the correct prediction with a

negated modal (i.e. "You can't go outside" was often interpreted as "Johnny went outside"), indicating some early competition in affirmative and negated reasoning. Even at an early age, children know how to make concrete, dichotomous decisions when modal verbs are used. (See also Bliss, 1988; Green, 1979; Hirst & Weil, 1982; Moore, Bryant & Furrow, 1989; Moore, Pure & Furrow, 1990.)

Adults also reason proficiently, in some cases facilitated by the use of modal verbs. In a Wason card selection task, participants made more errors with traditional if-then logic trials, when they were phrased as "if *p* then *q*" compared to those phrased with the deontic modal *must*, as in "if *p* then *must q*" (Cheng & Holyoak, 1985; see also Cosmides, 1989 and Girotto, Light, & Colbourn, 1988). These studies are offered as evidence for pragmatic constraints influencing decisions more than syntactic constraints. Thus, it is crucial to use natural stimuli one may hear in everyday conversation. This result also points to an amplification of certainty when *must* is used as compared with non-modal sentences.

These studies are informative because they provide insights into accuracy, but there remains a large gap between these and recent findings in decision-making and incremental, online speech perception. McKinstry, Dale & Spivey (2008) used computer-mouse tracking to study people's responses to phrases varying in degree of truth. They found that as uncertainty increases in a statement, deviations from a straight computer-mouse trajectory show up as a spatial attraction toward the competing response. For example, "You should brush your teeth everyday", rated as being completely true, showed a direct path to the "yes" response (with little curvature toward the "no" response), whereas the ambiguous phrase "Murder is sometimes justifiable" elicited substantially curved computer-mouse trajectories. Thus, making a decision about truth appears to be a process that unfolds over time, and is not made in terms of absolutes. Measurements of an ongoing response that begins while stimulus comprehension is still in progress can thus reveal subtleties in processing and brief considerations of alternative responses, which are not revealed in identification data (see Magnuson, 2005).

Some have referred to modal verb use as being indicative of degree of certainty, where deontic modal uses range from uncertain to certain (Close, 1975, p. 273). However, this hypothesis has not been tested perceptually, and it is unknown when or how this uncertainty is dealt with when a discrete response is required. Eyetracking can reveal probabilistic activations of two or more possible responses

unfolding over time, as eye fixations are closely time-locked to speech input. Studies that use this method have yielded much evidence for the idea that speech perception is incrementally processed (Allopenna, Magnuson & Tanenhaus, 1998; Altmann & Kamide, 1999; Spivey, Tanenhaus, Eberhard & Sedivy, 2002; Knoeferle, Crocker, Scheepers & Pickering, 2005), and that context is integrated continuously as a statement is unfolding (Sedivy, Tanenhaus, Chambers & Carlson, 1999). Thus, if modal verbs are simply contextual cues as to the intended strength of certainty, the process of coming to agreement or disagreement with a modal phrase should be quantifiable in terms of proportions of fixations over time.

Given that modal verbs presumably mitigate the truth of statements (Frawley, 1992), and that decisions about the truth appear to be probabilistically weighted according to degree of truth (McKinsty et al., 2008) it is reasonable to predict that different modal verbs will exert different forces on agreement dynamics. We predicted that *should* and *must* exert their influences differently, allowing for a temporary phase of considering an incorrect response. That is, the modal verb *should* may more readily allow the parallel consideration of both agreeable and disagreeable mental models of a given statement (or perceptual simulations of the possible states of affairs). If this uncertainty follows the Close (1975) scale, *should* will trigger more eye movements to the competing response than *must* because *should* is marking greater uncertainty.

The process of agreeing with a statement in an abstract sense may be made in terms of committing to reality, while simultaneously considering alternate, possible states of reality. We propose that predicates with modal verbs are processed as contingent statements, which are gradually resolved into one of the mental models, as they are being perceived.

The alternative hypothesis neglects interstitial stages of processing, and would be consistent with traditional linguistic definitions of modality. These grades of certainty hint at a probabilistic output, but make no predictions of competition being an intermediary mechanism in this decision-making process. Specifically, this would appear as a weakening or strengthening of looks to a target, with random looks to a competitor. This would be in line with certainty being a component, but the lack of competition would also predict reaction times to *must* statements would be faster than those to *should*. A lack of differences in looks to the competing response would also indicate a decision in terms of absolutes. In this case, the outcome could be modeled with traditional logical operations, which incorporate some degree of truth and/or certainty.

Methods

Participants. A total of 39 participants were run, 20 in the Should condition, and 19 in the Must condition. Each was right-handed or ambidextrous, native English speaking, and had normal or corrected to normal vision. Participants received course credit for participation. The study adhered

to the university's IRB standards, and included a debriefing session.

Stimuli. Thirty-one sentences were created, once using the modal *should* and then replicated for *must*, yielding 62 sentences. All were written in the second person (directed at the reader), for instance, "You must/should brush your teeth everyday" where 15 of each set were unambiguously agreeable, and 16 were unambiguously disagreeable as in "You must/should eat from a dirty plate". All stimuli are available from the first author upon request. A male speaker recorded these sentences in a quiet room multiple times, and each recording was selected on the basis of being similar and least variant in pitch, loudness and timbre. Silences were cut off the beginning and end of the recording at zero crossings to avoid audible pops, so there was no silence at the beginning or end of each file. All files were amplitude normalized. Final wav files were sampled at 44kHz in stereo sound, to be presented binaurally over headphones.

Text versions of the stimuli were rated by 100 undergraduate students at UC Merced. Survey participants were excluded from eligibility to participate in the main experiment. Each rated the sentences on a Likert scale of 1 to 7 on agreement-disagreement. For Disagree statements, 7 was the most disagreement, and 1 was the least disagreement, and for Agreement statements 7 was most agreement, and 1, least. All stimuli were on the high end of the scale for *must* (Agree: $M=6.1$, $SD=1.1$; Disagree: $M=6.1$, $SD=1.7$) and *should* (Agree: $M=6.1$, $SD=0.9$; Disagree: $M=5.7$, $SD=1.5$). Differences between Agree and Disagree were insignificant for each condition ($p>.1$).

Procedure. A screen displayed Agree or Disagree response boxes on the right and left side of the screen, with stimulus text presented in the middle approximately 200 pixels below the choices. In a drift correction before each trial, participants were required to press the spacebar on a keyboard while looking at a marker on the screen. This marker also functioned as a fixation point, to control eye position at the beginning of each trial. The left and right response boxes were 250x250 pixels, and contained the word "Agree" or "Disagree," with their relative positions randomized across trials. Simultaneous with the onset of the written sentence, a matching auditory file was played. The spoken instructions that were given to the participants in advance of the experiment informed them about the kinds of sentences they would encounter, and asked them to respond with either the left or right arrow key to indicate which side of the screen contained their chosen Agree/Disagree response. Modal verb (*should/must*) was a between subjects variable to disguise the key manipulation of the study.

Participants wore a head mounted Eyelink II eyetracker, sampling at 500 Hz. In addition to eyemovements, the response, reaction time, and side of screen were collected.

Results

Responses were extremely accurate, with an average of 30 correct responses ($SD: 2.6$) and with 14 of the total 39

participants responding accurately to all stimuli. The Should and Must accuracy averages were similar, with Should at about 95% correct (*SD*: 1%) and Must at about 96% correct (*SD*: 5%). Incorrect responses are excluded from all analyses.

The final response occurred at an average of 2534ms (*SD*: 799ms). Reaction times by condition are presented with standard deviations in parentheses here in Table 1.

Table 1: Reaction times by condition

	Positive	Negative
Should	2503ms (542ms)	2454ms (565ms)
Must	2645ms (797ms)	2535ms (1178ms)

A 2x2 mixed ANOVA was conducted to explore differences, with Should versus Must as a between-subjects variable, and Positive versus Negative as a within-subjects variable. Should and Must did not differ as well as Positive and Negative showing no difference or any interaction (all tests: $p > .1$). To further explore if one condition may contain differences, separate paired-samples t-tests were conducted. Both conditions contained no significant difference between Positive and Negative statements ($p > .1$).

Eyetracking. One participant in the Should condition was excluded from all eyetracking analyses due to experimenter error. Fixations were coded in a binary fashion over time to 1 of 4 areas of the screen: target, competitor, text, or blank regions. Each port was extended 100 pixels in each direction to allow for error in the track. Graphs of the time course for each condition can be found in Figure 1.

Overall proportions of fixations were subjected to independent samples t-tests, to examine differences between looks to the target, competitor, text and blank regions of screen between the two conditions, for both Positive and Negative stimuli. In a comparison of Should and Must for Positive trials only, looks to target, text and blank regions of screen did not reliably differ ($p > .1$). Looks to the competitor reliably differed ($t(36)=2.4$, $p=.02$), with more fixations to the competitor for the Should condition (M : 0.097, SD : 0.044) than for Must (M : 0.064, SD : 0.065). A similar trend is seen for Negative trials, with fixations to the target and blank regions no different from one another ($p > .1$), fixations to the competitor greater in the Should condition ($t(36)=2.9$, $p=.006$), and fixations to text marginally significant with more looks to text for Must ($t(36)=-1.8$, $p=.08$). Thus, the prominence of the competitor is greater for *should* than for *must* regardless of response type, but not at the cost of looking to the target in particular.

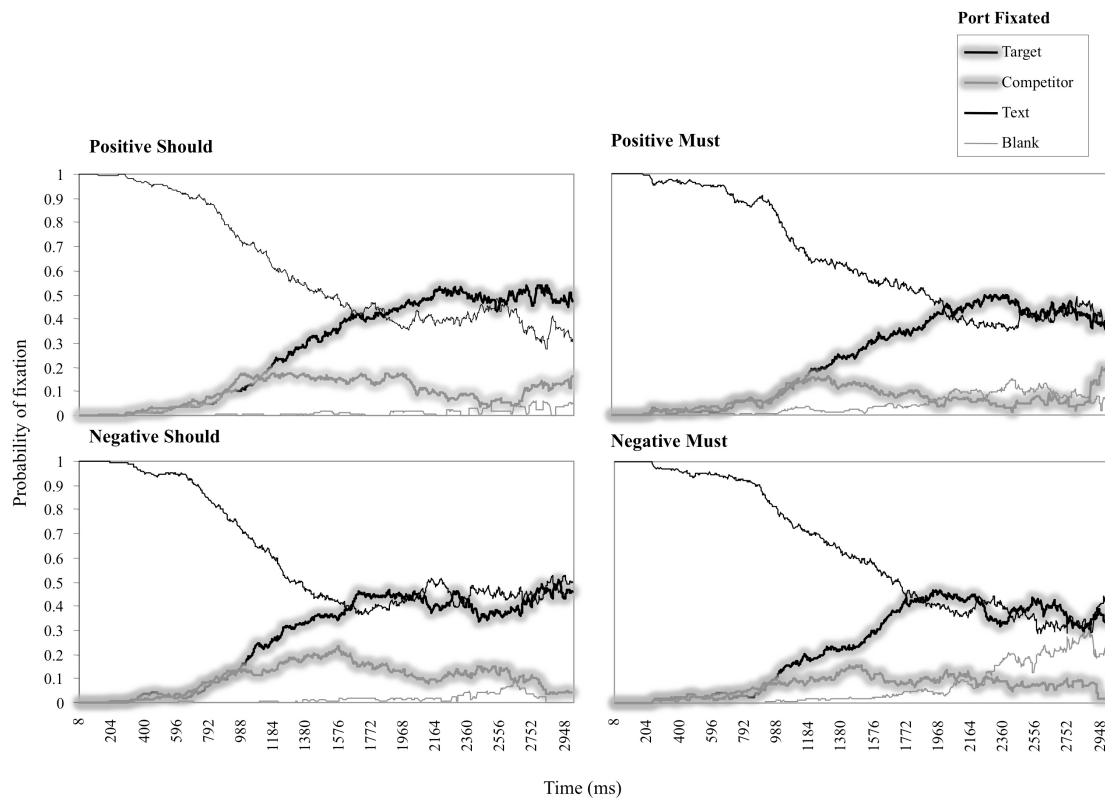


Figure 1: Timecourse by condition – Each panel represents one of four conditions, where Must and Should are between-subject conditions, and Negative (where the target is Disagreement) versus Positive (target Agreement) stimuli. Each line represents the probability of fixating on that particular region of the screen at that moment in time.

To test whether looks to the competitor in the Must condition were negligible or not, a one-sample t-test against a value of zero was performed. Positive Must was significant for the competitor against a value of zero ($t(18)=7.61$, $p<.001$) as was Negative Must ($t(18)=9.68$, $p<.001$). This analysis shows that there are still some looks to the competitor and thus, that processing statements with *must* is not likely to be an all or nothing process.

Because these tests examine positive and negative stimuli separately, it can be useful to look at the differences between Positive and Negative trial fixations over time. There may be a window of time during which there are more fixations to a certain region when Positive stimuli are heard, and another window during which a region is more prominent for a Negative stimulus. By pooling all time in previous analyses, there is a risk of averaging and washing out these differences. Graphically we can examine this by subtracting fixations in Negative trials from the Positive for each time step. Thus, the y-axis becomes a kind of valence scale, where 0 indicates no difference between looks for Positive and Negative stimuli. Positive y means a bias toward that particular port in Positive trials, and negative y is bias during Negative trials. Figure 2 shows that in the Should condition, between 1244ms and 2200ms there is a bias toward looking at the Negative Target, and a bias during this same time window toward looking at Positive Text. These differences were also mapped out for the Must condition, but were uninformative and will not be investigated further here.

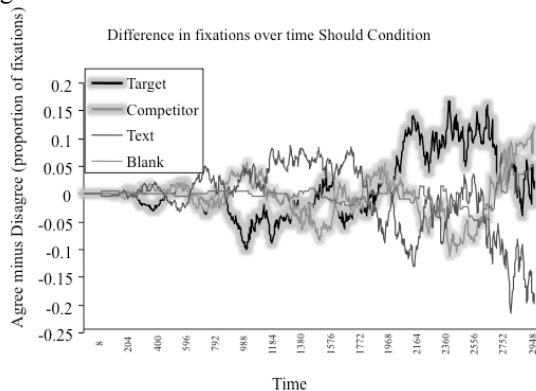


Figure 2: Difference in average fixations over time – This graph is constructed by subtracting average fixations on Disagree trials from Agree trials. Divergence from no difference was observed at 900ms through 1799ms for Text and Target areas of the screen.

Time was binned into Beginning (0-899ms), Middle (900-1799ms) and Final (1800ms-3000ms). These time bins were analyzed in a similar way to the previous fixation analyses, where average proportions of fixations to each port were output for each subject in each condition, and then compared across the Positive / Negative dimension in a paired samples t-test. As predicted, there were no differences in looks to any regions of the screen in the Beginning time window or the Final window ($p>.1$).

Fixations to the Target in the Middle time window were greater for the Negative stimuli (Positive: $M=0.28$, $SD=0.11$; Negative: $M=0.322$, $SD=0.14$; $t(18)=2.17$, $p=.043$). Other ports revealed no differences ($p>.1$).

Discussion

For the Should condition, the increased looks to the competitor seemed to indicate that the competing response is more active throughout the trial than in the Must condition. A lack of differences in looks to the target where the competitor is prominent indicates that this increase is not competition, but rather an increase that comes at the cost of no particular region of the screen. Although there are very few looks to the competitor in the Must condition, a test against a value of 0 revealed that there are still some occasional looks here. However, because of the random assignment of side of response, we cannot rule out the possibility that this was due to scanning for the location of the correct response, instead of a brief consideration of responding in this way. Reaction times were uninformative, but further investigation of the middle of the trial revealed that the correct target Disagree is fixated on more, showing an earlier rise in looks to the target on Negative trials in the Should condition. For example, when the stimulus was “You should eat from a dirty plate”, the participant fixated on the correct Disagree port earlier than with a Positive stimulus such as “You should study hard for exams”.

Previous work has made bold claims for a domain-specific, innate, deontic-reasoning module (Cummins, 1996; see also Gigerenzer & Hug, 1992). Also, while not making direct comparisons between different modal verbs, Traxler, Sanford, Aked & Moxey (1997) indicate that one of their assumptions is that all modal verbs carry a possibility marker. As it has been shown here though, deontic reasoning takes place as the speech is being heard, as revealed by eye fixations. Because this decision is not solely driven by auditory information, but must rely on visual input and rapid sharing with areas that drive eye movements, it is unlikely that deontic-reasoning is domain-specific. Further, since the decision is being made as semantic and syntactic information are still coming in, it would be difficult to integrate a specific module for deontic reasoning without having continuous access to semantic and pragmatic constraints as this information arrives, which would be in line with constraint-based theories of sentence processing (e.g., Tanenhaus & Trueswell, 1995).

The semantic theory of possible worlds (Hintikka, 1975) fits with these results, where the set of possible worlds could be compared to output nodes in a dynamic neural network. Upon hearing a sentence with the word “should”, the set of output nodes would have some activation for semantic features of the proposition, some true (partial activation above some threshold), and some false (below threshold activation). For example “You should eat from a dirty plate” would perhaps simulate a dirty plate, a clean plate (an alternative based on experience), motor plans of eating motions, etc. As time moves forward, although

“dirty plate” is explicitly stated, the activation for the dirty plate node would gradually fade while the clean plate node ramps up. Thus, although it is possible to eat from a dirty plate, and there may be some temporary fleeting time in which this is true in some possible world, we gradually resolve into a set of possible worlds or nodes based on previous experience in our world. While we do not propose any mechanism of how differences in certainty might be learned, it is possible to create a kind of possible worlds network with a simulation of simple visual, auditory, sensory feature nodes in combination with a simple recurrent word learning network (Anderson, Huette, Matlock, & Spivey, in press; Howell, Jankowicz & Becker, 2005).

These results reveal that agreement with obvious, unambiguous, everyday standards of conduct can be strongly affected by the choice of modal verb. Highly unambiguous sentence processing unfolds over time, and is susceptible to the influence of subtle variation in verbal modality. It is likely that these intermediate stages of processing are crucial to how one perceives, and makes decisions when these modal auxiliaries are used in everyday discourse, and an integration of compatible linguistic and psycholinguistic theories will be necessary for resolving further debates on the nature of reasoning where modal verbs are involved.

References

- Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419-439.
- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247-264.
- Anderson, S.E., Huette, S., Matlock, T., & Spivey, M.J. (in press). On the temporal dynamics of negated perceptual simulations. In F. Parrill, V. Tobin, & M. Turner (Eds.), *Meaning, form, & body*, (pp. 1-20). Stanford, CSLI.
- Bliss, L.S. (1988). Modal usage by pre-school children. *Journal of Applied Developmental Psychology*, 9, 253-261.
- Byrnes, J.P., & Duff, M.A. (1989). Young children's comprehension of modal expressions. *Cognitive Development*, 4(4), 369-387.
- Cheng, P., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416.
- Close, R.A. (1975). A reference grammar for students of English. Essex: Longman Group.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31(3), 187-276.
- Cummins, D.D. (1996). Evidence for the innateness of deontic reasoning. *Mind & Language*, 11(2), 160-190.
- Divers, J. (2002). Possible Worlds. London: Routledge.
- Frawley, W. (1992). *Linguistic Semantics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition*, 43, 127-171.
- Giroto, V., Light, P. H., & Colbourn, C. J. (1988). Pragmatic schemas and conditional reasoning in children. *Quarterly Journal of Experimental Psychology*, 40A, 342-357.
- Green, M. (1979). The Developmental Relation between Cognitive Stage and the Comprehension of Speaker Uncertainty. *Child Development*, 50(3), 666-674.
- Halpern, J.Y., & Moses, Y. (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54, 319-379.
- Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4(3), 475-484.
- Hirst, W., & Weil, J. (1982). Acquisition of epistemic and deontic meaning of modals. *Journal of Child Language*, 9, 659-666.
- Howell, S. R., Jankowicz, D., & Becker, S. 2005. A Model of Grounded Language Acquisition: Sensorimotor Features Improve Lexical and Grammatical Learning. *Journal of Memory and Language*, 53, 258-276.
- Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering, M.J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95(1), 95-127.
- Magnuson, J. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences of the USA*, 102(29), 9995-9996.
- Marslen-Wilson, W.D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198), 226-228.
- McKinstry, C., Dale, R., & Spivey, M.J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, 19(1), 22-24.
- Moore, C., Bryant, D., & Furrow, D. (1989). Mental Terms and the Development of Certainty. *Child Development*, 60(1), 167-171.
- Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, 61, 722-730.
- Nuyts, J. (2006). Modality: Overview and linguistic issues. In W. Frawley (Eds.), *The expression of modality*, (pp. 1-26). Berlin, Mouton de Gruyter.
- Sedivy, J.E., Tanenhaus, M.K., Chambers, C.G. & Carlson, G.N. (1999). Achieving incremental interpretation through contextual representation: Evidence from the processing of adjectives. *Cognition*, 71, 109-147.
- Spivey, M.J. Tanenhaus, M.K., Eberhard, K.M. & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447-481.

- Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- Tanenhaus, M. K. & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of Perception and Cognition*. (pp.217-262). San Diego: Academic Press.
- Traxler, M.J., Sanford, A.J., Ake, J.P., & Moxey, L.M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 88-101.