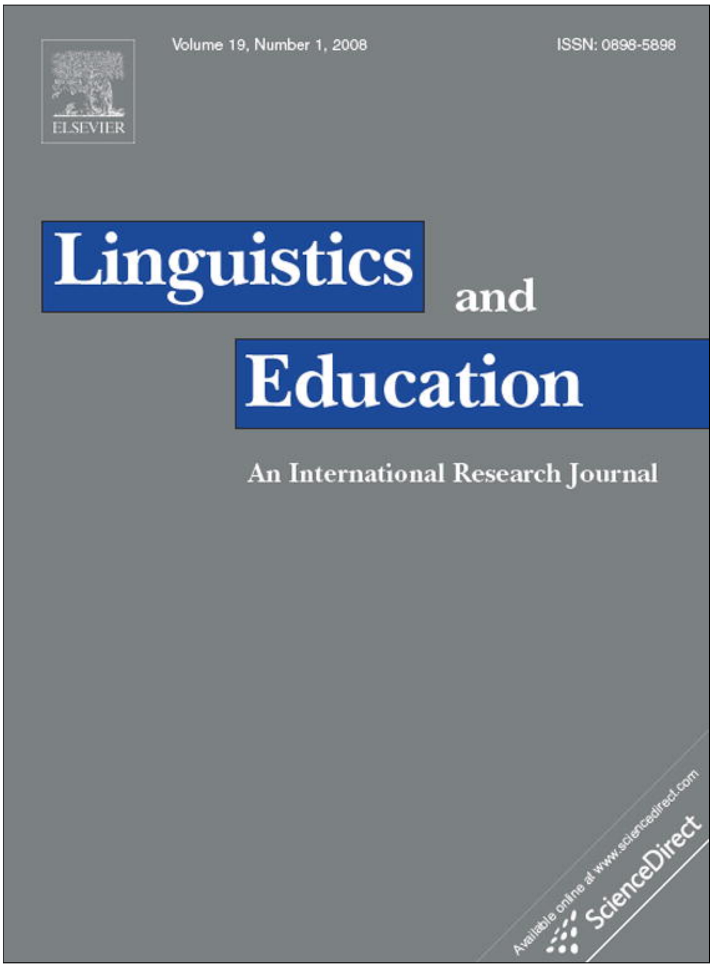
Volume 19, Number 1, 2008
ISSN: 0898-5898
ELSEVIER
Linguistics and Education
An International Research Journal
Available online at www.sciencedirect.com
ScienceDirect

Available online at www.sciencedirect.com
ScienceDirect
Linguistics and Education 19 (2008) 56–69
Linguistics and Education
www.elsevier.com/locate/linged

What if? Conditionals in educational registers

Max M. Louwerse[a,*], Scott A. Crossley[b], Patrick Jeuniaux[a]

[a] *University of Memphis, TN, United States*
[b] *Mississippi State University, MS, United States*

Abstract

Many corpus linguistic studies have investigated classification of texts into genres and registers, but relatively few of these studies have looked at linguistic features in educational registers. From a pedagogical perspective it is important to determine whether certain linguistic features behave differently across registers within particular disciplines. The current study investigates conditionals, linguistic features that have extensively been studied in psychology, philosophy and education and can provide useful information on student reasoning and misconceptions. A series of corpus linguistic studies show that conditionals are more frequently used in a discipline like physics, regardless of the language mode (monolog or dialog), educational environment (tutoring or non-tutoring), formality (formal–informal) or interface (human–human, human–computer). Furthermore, evidence was found that students who are more proficient in physics use a higher number of conditionals than less proficient physics students. These findings have consequences for corpus linguistic studies on registers, for general pedagogy, and for physics pedagogy in particular.
Published by Elsevier Inc.

Keywords: Conditionals; Genre; Register; Corpus linguistics; Discourse; Conceptual physics; Intelligent tutoring systems

Various corpus linguistic studies have investigated linguistic features for a variety of situational characteristics in text and discourse according to their purpose, setting, interactiveness or mode (see Biber, Conrad, & Reppen, 1998; Crossley & Louwerse, 2007; McEnery & Wilson, 1996). These varieties are commonly called 'registers' (Biber, 1988). Generally, these studies divide registers into categories like narrative versus expository, or written versus spoken discourse (Louwerse, McCarthy, McNamara, & Graesser, 2004).

Such classifications can shed light on general language use, but give little insight in differences across educational registers. That is, these corpus linguistic studies do not address the fact that different academic communities have distinct discourses with unique linguistic features (Hyland, 2000). At the same time, identifying linguistic features more common in one discipline than in others can have important implications for a general understanding of the discipline, for instance in terms of the interactions between writers and their readers in published academic writing (Hyland, 2000). Linguistic insights such as these can influence the teaching of a discipline. They might also aid in student assessment, whereby the student language is compared to the language appropriate for a discipline. In such a case, student ability in a discipline could be measured in part by the linguistic features employed by the student.

For reasons to be discussed below, this paper focuses on the discipline of physics in order to investigate one such linguistic feature, that of conditionals. Conditionals are those linguistic expressions that mark a condition

* Corresponding author at: Department of Psychology, Institute for Intelligent Systems, University of Memphis, Psychology Building, Memphis, TN 38152, United States. Tel.: +1 901 678 2143; fax: +1 901 678 2579.
E-mail address: mlouwers@memphis.edu (M.M. Louwerse).

0898-5898/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.linged.2008.01.001

making the accompanying expression contingent on the truth or plausibility of that condition. That is, speakers express, define, and/or confirm under a certain condition allowing them to explicitly reason or to explicitly consider hypothetical scenarios. Conditionals are often marked by *if-then* sentences and are frequent in many registers of language use. Distinguishing their use in academic genres would have important implications for both knowledge assessment and disciplinary pedagogy. One of the motivations to choose conditionals is their importance in scholarly discussions in a variety of disciplines. In psychology, conditionals have been studied to understand aspects of human reasoning (Johnson-Laird, 1983). In philosophy, conditionals shed light on truth conditions (Strawson, 1986). In linguistics, conditionals have been studied to determine the interdependence of the form and interpretation (Dancygier, 1999). Our personal motivation behind investigating linguistic features is driven by our work in intelligent tutoring systems. As outlined below, these intelligent systems ask student questions in a particular discipline, analyze student input, and, depending on the student input, choose pedagogically and conversationally appropriate dialog moves. In the computational understanding and production of language, identifying idiosyncrasies in language use related to a discipline is essential. In addition, the use of conditionals might help these systems in detecting misconceptions, because the reasoning process is made explicit in conditional structures. Their use also helps in creating common ground, something all conversational partners (including intelligent tutoring systems) need to establish.

In sum, this paper explores the use of conditionals combining both an educational and corpus linguistic perspective in order to shed light on language use in teaching. We will do this by first providing an overview of the use of conditionals in language and outlining the importance of them for education. This will be followed by a series of corpus linguistic studies comparing the use of conditionals according to registers and domain.

1. Conditionals in language use

Conditionals generally take the form of *if-then* sentences, whereby the subordinate clause (the protasis, p) carries the conditional conjunction (*if*) and is dependent on the main clause (the apodosis, q). The order of the protasis and apodosis in the linguistic expression can vary, dependent on backward or forward relations (Louwerse, 2002). Most studies on conditionals are concerned with the truth value of the propositions (Barrouillet, Grosset, & Lecas, 2000; Johnson-Laird, 1983). For formal logic this is of course important, but in the current study truth values are not as relevant, because the focus lies on how conditionals work rather than their propositional content. Because this paper will be dealing with discourse rather than propositional representations in a formal logic, we will ignore a truth functionalist view on conditionals and consider any utterance with the format of an *if-then* sequence to be a conditional sentence. These sequences will be the central topic in our corpus linguistic study.

2. Conditionals in education

Why would a corpus linguistic study of conditionals be of any interest to the education community? First of all, linguistic features in general are able to shed light on language use in education as well as teaching particular disciplines. The study of student uses of conditionals allows an educator to gain an understanding of those students' lines of reasoning. Consider for instance example (1), taken from the *Why2 Corpus of Physics Tutoring* (see Graesser, VanLehn, Rose, Jordan, & Harter, 2001) tutoring dialog. The student responds to the following physics problem: "You are given two containers in a weightless condition in outer space. One is filled with feathers, the other with lead. How could you determine which one was heavier?"

(1) *Student*: You could apply a force to each of the containers.
Tutor: How would you apply a force to determine which one was heavier?
Student: If we push both the containers with the same force, the lighter container will accelerate faster since acceleration is inversely proportional to mass.

During the instructional process the student constructs answers to questions given various states and constraints. The use of conditionals provides rationales, depth, and evidence of reasoning or the reasoning process. This can be highly informative for educators of all types and can help them to tailor learning to the student.

An additional reason related to the understanding of student reasoning concerns misconceptions: intuitive ideas in a domain which are at odds with the formalized body of knowledge in this domain (VanLehn, 1990). Particularly in the field of physics, these misconceptions have been carefully identified to compare and contrast physics concepts to commonsense alternatives (Halloun & Hestenes, 1985; Hestenes, Wells, & Swackhamer, 1992). Hypothetical reasoning easily reveals misconceptions. Take for instance example (2) from the same container problem in the tutoring transcripts.

(2) If the containers were in outer space, there is no real way to determine which one has more mass.

The implication exhibited by the student is a misconception: the intuitive idea of less gravitational force being related to less mass is at odds with the formalized body of knowledge in this domain which states that mass and weight – the latter indeed being affected by gravitational force – are different from each other. Since there are multiple ways to determine which of the containers is heavier in a weightless condition, we can assume that the student in example (2) demonstrated a fundamental misunderstanding about weightlessness in outer space, unlike the student in example (1), who selected a method that is sound. With this information in hand, tailored instruction can correct the misconception and modify the learning to address the problem. For instance, a tutor could address the fact that the containers in space would have the same resistance to a change in motion as they would have on earth and that, therefore, the ways to determine mass on earth would also apply in outer space. Example (2) points out how potent conditional statements exist in an educational setting and how diagnostic they can be in instruction.

A third reason conditionals are of interest to the educational community has to do with the contrast that has frequently been made between shallow and deep knowledge (Bloom, 1956; Bransford, Goldman, & Vye, 1991; Chi, de Leeuw, Chiu, & LaVancher, 1994; Otero, Leon, & Graesser, 2002). Shallow knowledge relates to facts and definitions, or the explicit concepts and ideas in the learning material. Deep knowledge on the other hand consists of explanations of the learning material. For a deep understanding of the material, and for planning and reasoning, shallow knowledge does not suffice. Conditionals motivate deep knowledge, and the use of them is evidence for this deeper knowledge. Take for instance the following two physics questions:

(3) Does velocity relate to speed?
(4) If a runner runs in a straight line at constant speed and throws a pumpkin straight up, where will the pumpkin land?

According to Graesser, Person, and Huber (1992) the question in (3) can be identified as a Yes/No question, which typically yields short answers, whereas the question in (4) is identified as an Interpretation question, which typically yields a long answer. That is, the question-content categories vary in the length of the expected answers, with the questions inviting short answers placing fewer demands on the answerer than those requiring long answers. Short-answer questions are therefore classified as shallow-comprehension questions and long-answer questions as deep-comprehension questions (Graesser & Person, 1994; Graesser et al., 1992). Graesser and Person (1994) found a difference in student performance depending on whether they asked shallow or deep questions, with deep comprehension questions having a significant positive correlation with examination scores. From their study, it could be argued that questions involving conditionals tend to yield more elaborate answers, and can therefore be classified as deep comprehension questions.

One final reason why a corpus linguistic study of conditionals is useful has to do with intelligent tutoring systems. Intelligent conversational tutoring systems are computational tutors that communicate with students using natural language (as compared to mathematics or formulaic language) and additional information like pictures or interactive simulations. These systems ask questions of the students, interpret the student answers, and answer student questions. They have conversations with the student similar to those with a human tutor (see Graesser et al., 2001). Since the 1980s, intelligent tutoring systems have proven to be very successful. From a theoretical, academic perspective these systems have brought together the disciplines of artificial intelligence, computer science, cognitive science, psychology and linguistics. But these systems have also proven to be successful in a range of learning environments, including geometry, algebra, physics and electronics. Indeed, the generations of systems currently developed have shown to provide considerable learning gains in student test scores up to 1 standard deviation compared with test scores for untutored students learning the same content in a classroom (Corbett, Anderson, Graesser, Koedinger, & van Lehn, 1999). In fact, unaccomplished human tutors have been reported to enhance learning with an effect size of 0.4 standard deviation on test scores, whereas intelligent tutoring systems produce effect sizes of approximately 1.0 standard deviation (Corbett, 2001).

The latest generation of intelligent tutoring systems use conversational interfaces. Students can interact with the conversational system using written or spoken dialog and the system responds in writing or in speech. For computational systems that deal with natural language processing (NLP), an understanding of specific language features in different registers is important (Jurafsky & Martin, 2000). Furthermore, the hypothetical situations that can be constructed using conditionals require a careful semantic analysis. For multiple reasons (theoretical and applied linguistics, language teaching, teaching, understanding student reasoning, deep knowledge, and intelligent tutoring systems), analyzing the frequency of conditionals in different text registers is an important endeavor.

3. The use of conditionals in different registers

What is the general role of conditionals in language use? Corpus linguistic studies have explored this question by calculating their frequency in various registers in a variety of circumstances with obvious varying situational characteristics. Biber (1988) for instance conducted an extensive corpus linguistic study and investigated the similarities and differences between various language settings. He compared 67 linguistic features in 481 corpora across 23 written and spoken registers and identified 6 underlying dimensions of variation in English: involved versus informational production of discourse, narrative versus non-narrative concerns, explicit versus situated reference, overt expression of persuasion, abstract versus non-abstract information and on-line information elaboration. The frequency of conditionals in each of these dimensions derived from the 23 registers clarifies the difference in conditional use across in registers. Table 1 presents these frequencies derived from Biber (1988).

For exploratory purposes Table 1 gives a general overview of conditionals in language use. Based on these average frequencies, conditionals tend to be more frequent in discourse that deals with involved production (like face-to-face

Table 1
Comparison frequency conditionals on six dimensions

Dimension label			Mean (S.D.)
1	Involved production	Face-to-face conversations, interviews, personal letters, spontaneous speeches	3.53 (2.40)
	Informational production	Academic prose, adventure fiction, biographies, broadcasts, general fiction, hobbies, humor, mystery fiction, official documents, popular lore, prepared speeches, press editorials, press reportage, press reviews, professional letters, religion, romantic fiction, science fiction, telephone conversations	2.21 (1.88)
2	Narrative concerns	Adventure fiction, general fiction, mystery fiction, science fiction	2.65 (1.93)
	Non-narrative concerns	Academic prose, biographies, broadcasts, face-to-face conversations, hobbies, humor, interviews, official documents, personal letters, popular lore, prepared speeches, press editorials, press reportage, press reviews, professional letters, religion, romantic fiction, spontaneous speeches, telephone conversations	2.40 (1.98)
3	Explicit reference	Academic prose, biographies, hobbies, popular lore, prepared speeches, press editorials, press reviews, professional letters, religion, spontaneous speeches	2.10 (1.96)
	Situation-dependent reference	Adventure fiction, broadcasts, face-to-face conversations, general fiction, humor, interviews, mystery fiction, official documents, personal letters, press reportage, romantic fiction, science fiction, telephone conversations	2.70 (1.98)
4	Overt expression of persuasion	Adventure fiction, broadcasts, general fiction, hobbies, interviews, mystery fiction, official documents, personal letters, prepared speeches, press editorials, religion, romantic fiction, spontaneous speeches, telephone conversations	2.72 (2.11)
	No or covert expression of persuasion	Academic prose, biographies, face-to-face conversations, humor, popular lore, press reportage, press reviews, professional letters, science fiction	2.00 (1.77)
5	Abstract information	Academic prose, hobbies, official documents, popular lore, press editorials, press reportage, press reviews, professional letters, religion	1.89 (1.96)
	Non-abstract information	Adventure fiction, biographies, broadcasts, face-to-face conversations, general fiction, humor, interviews, mystery fiction, personal letters, prepared speeches, romantic fiction, science fiction, spontaneous speeches, telephone conversations	2.79 (1.98)
6	On-line informational elaboration	Academic prose, face-to-face conversations, interviews, prepared speeches, press editorials, professional letters, religion, spontaneous speeches	2.74 (2.03)
	No on-line informational elaboration	Adventure fiction, biographies, broadcasts, general fiction, hobbies, humor, mystery fiction, official documents, personal letters, popular lore, press reportage, press reviews, romantic fiction, science fiction, telephone conversations	2.28 (1.95)

Data derived from Biber's (1988).

conversations) than in discourse that deals with informational production (like academic prose). Discourse that involves narrative concerns (like fiction) has only a slightly higher number of conditionals than the non-narrative discourse (like professional letters). The difference is more distinct in discourse dealing with situation-dependent reference (like humor) that has more conditionals than discourse that uses explicit reference (like speeches). Similarly, in discourse that uses overt expression of persuasion (like telephone conversations) conditionals are higher than discourse with no or covert expressions of persuasion (like biographies). Non-abstract information (fiction) generally uses a higher number of conditionals than discourse that uses abstract information (academic prose). Finally, on-line informational elaboration (like interviews) tends to use more conditionals than discourse that does not use on-line elaboration (like personal letters).

In sum, corpora that deal with involved production, that are non-abstract and narrative, and that use situation-dependent reference, overt persuasion and on-line informational elaborations, tend to have the highest number of conditionals. This means that in registers like personal letters, face-to-face conversations, spontaneous speeches, interviews and fiction, a higher number of conditionals can be expected; in press reports, press reviews, popular lore and academic prose a lower number can be expected.

In addition to studies investigating general discourse registers like spoken and written discourse, some studies have investigated the use of conditionals in domain specific discourse. Mead and Henderson (1983) for instance found that in economic text conditionals were prevalent, but that they did not always relate to the concepts of the text, while Ferguson (2001) found that there was significant variation in conditional use between written and spoken medical discourse with conditionals as politeness strategies common in consultation, but uncommon in doctor-to-doctor written communication. Lastly, Horsella and Sindermann (1992) looked at both spoken and written discourse in the social and medical sciences and found that while deductive and inductive reasoning was common, it was often conditionals that provided semantic links between the premise and the conclusion.

Although the frequency of conditionals in the various dimensions has been presented here for exploratory purposes, it does raise an important question for educational materials like textbooks and pedagogical methods such as tutoring and classroom instruction. It seems that most educational registers can be classified into the dimensions that are not expected to have a high number of conditionals. Because of the contrast between the usefulness of conditionals in education and the expected low number of conditionals, a further investigation into the use of conditionals in educational disciplines becomes more important. A series of corpus linguistic studies using a variety of additional educational registers serves this purpose.

4. Corpus linguistic studies

Before we conduct any frequency analyses, it is worth revisiting the definition of conditionals for our purposes. So far, we have defined conditionals as *if-then* sentences. However, discourse uses a wider variety of conditionals markers. Examples from tutoring transcripts are given in (5).

(5) a. *When* the rock has been released by the man, is there still a force due to him? [If the rock has been released by the man, then is there still a force due to him?]

b. Fine, *suppose* its initial velocity when released by man was v, what will be its velocity when it reaches the same level as released? [Fine, if its initial velocity when released by man was v, then what will be its velocity when it reaches the same level as released?]

c. Alright, *if* the rock falls through this distance, what will be its final velocity? [Alright, if the rock falls through this distance, then what will be its final velocity?] Now, *assuming that* they are both traveling the same distance. They will both have the same velocity, so over time their total velocity will always be equal. [Now, if they are both traveling the same distance, then they will both have the same velocity, so over time their total velocity will always be equal.]Well it is the time from release to the time it is caught. The change in velocity is the same for both the rise and the fall, and the acceleration is the same for the rise and the fall, *therefore* it must be equal for the rise and the fall. Well it is the time from release to the time it is caught. If the change in velocity is the same for both the rise and the fall, and the acceleration is the same for the rise and the fa*ll, then i*t must be equal for the rise and the fall].

For the current study, these expressions are considered conditionals if they can be replaced by *if-then* constructions without changing the meaning of a sentence. This substitution test is not uncommon in corpus linguistics (see Knott, 1996; Louwerse & Mitchell, 2003) and allows for the computational investigation of a large number of conditionals. Following this definition of conditionals, 106 markers synonymous (or contingently substitutable) with *if-then* were selected. These markers were derived from Roget's Thesaurus and are given in Table 2. Note that markers like *because*

Table 2
Conditionals (substitutions for *if-then* derived from thesaurus)

According as	Imaginable	Provided
Agreed	Imagine	Provided always
Allowing that	In any case	Provided that
As if	In case	Providing
As long as	In case that	Provisionally accept
As though	In either case	Provisionally admit
Assumable	In idea	Provisionally agree to
Assume	In Never-Neverland	Reckon
Assume for argument's sake	In the Abstract	Say
Assumed	In the contingency that	Say for the hell of it
Assuming that	In the event that	Stipulated
Assumptive	In theory	Stipulating
Am inclined to think	Infer	Subject to
By way of hypothesis	Inferred	Supposable
Conjecturable	It being provided	Suppose
Conjectured	Just in case	Suppose that
Considering	Let	Supposed
Daresay	Let's say that	Supposing
Deduce	On condition	Supposing that
Deemed	On condition that	Suppositive
Expect	On paper	Surmisable
Gather	On the assumption that	Surmise
Given	On the supposition that	Suspect
Given the fact that	Opine	Take as a precondition
Given that	Postulated	Take for
Granted	Postulational	Take it
Granted that	Prefigure	Take it as given
Granting	Premised	Taken as
Granting that	Premissable	Taken for granted
Hypothetically	Presumable	Theoretically
Ideally	Presume	Understand
If	Presumed	Understood
If ever	Presumptive	With the stipulation
If it should happen that	Presuppose	With the understanding
If only	Presupposing	With this proviso
If we assume that	Presurmise	

and *so* are not included because the implicature of the causal relation may be a conditional, but that the markers itself does not always mark conditionals.

Without claiming that this set is exhaustive, we assume that the 106 markers form a representative set of conditionals markers and their frequencies will give a reasonable picture of conditionals in corpora.

5. Study 1

Earlier, we argued that an investigation of conditionals in educational material is useful because it motivates and illustrates deep reasoning, reveals misconceptions, and can help intelligent tutoring systems. Because physics tutoring dialog can be expected to consist of deep reasoning questions as defined according to the Graesser et al. (1992) and the Graesser and Person (1994) taxonomies, we can predict that it contains a high number of conditionals compared to a general language use. To test this hypothesis we compared the overall frequency of conditionals in a corpus of physics tutoring dialog with both spoken and written language.

For the physics tutoring corpus, we used the *Why2 Corpus of Physics Tutoring* (see Graesser et al., 2001). This corpus consists of data from tutoring dialogs between expert human tutors and physics students. Experienced physics tutors tutored a total of 11 college students for about 3 h each on 5 physics problems. A student and tutor were seated in separate rooms and communicated using a computer screen and keyboard. This way of tutoring comes closest to what a computational tutor does. In the *Why2 Corpus of Physics Tutoring*, the tutor presented the student with a series of problems, the same for all students. The student solved the problem by submitting an essay, which was then evaluated by the tutor. At the same time, both the tutor and the student could exchange meta-tutoring information. The student could for instance ask for explanations and could make analogies.

As the comparison corpus we used the British National Corpus (BNC). The BNC is comprised of 100 million words and contains samples of modern written and spoken language from a wide range of sources. The written part includes texts from newspapers, journals, academic books and popular fiction. The spoken part includes informal conversation, as well as formal business or government meetings, as well as radio shows and phone-ins.

First, the corpora were cleaned up by removing any tagging and coding. All conditionals that matched the 106 markers mentioned in Table 2 were selected regardless of words homonymous to these conditional markers. Next, all selected conditionals were checked on a case-by-case basis to determine whether the if-then substitution was justified, i.e., matched the original if-then meaning. Finally, in order to adjust the raw frequency counts from texts of different lengths, a common normalization procedure was applied (see Biber et al., 1998): The frequency of conditionals was divided by the total number of words in the corpus and multiplied by 10,000, to norm the frequency to a basis per 10,000 words. Although this number is an arbitrary constant value, the number 10,000 was used to account for a normalized frequency that comes closest to the original frequency value. Similarly, to obtain the number of conditionals per turn in the case of dialogs, the frequency of conditionals was normalized to a percentage (per 100 turns basis).

In addition to the total frequency of conditionals, the relationship between the types (the number of separate vocabulary items) and tokens (the total number of occurrences of words) was calculated to account for lexical variation in the conditional use. But a simple division of types by tokens does not take into consideration the text length. To normalize these counts we used Herdan's (1966, pp. 75–77) proposal of dividing the logarithm of the number of types by the logarithm of the number of tokens.

Using the BNC is one of the best approximations of frequencies in modern language use. However, its sheer size makes it infeasible to check on a case-by-case basis whether the if-then substitution was justified. Therefore, the normalized frequencies are for non-verified conditionals and are hence an overestimation of the actual count. Results showed that of the 679 normalized conditionals found in the Why2 corpus, only 446 normalized cases were found in the BNC corpus ($\chi^2_{(1)} = 24.03, p = 0.001$). The type–token ratio showed a similar pattern, with more different conditionals for the Why2 corpus (0.42) than for the BNC corpus (0.30). In sum, the Why2 dialog physics tutoring corpus has a very high frequency of conditionals compared to the general frequency of words that are homonymous with conditionals markers. It can thus be concluded that compared to spoken and written language, computer mediated physics tutoring dialog has a higher number of conditionals.

6. Study 2

The finding in study 1 can be attributed to a variety of factors. They are presented next in the form of hypotheses:

1. The Why2 corpus is *dialog*. The corpus contains dialogs between tutors and students. Since dialogs are more conversational than monologs (or a variety of them, as in the BNC), they contain more imaginary situations and less common ground. Dialog participants need to establish common ground and this requires a higher frequency of conditions. Dialogs therefore consist of more conditionals than monologs.
2. The Why2 corpus is *tutoring discourse*. Tutoring sessions require more theoretical situations, as the tutor needs to give hypothetical examples as for instance in the container-in-space or the pumpkin-throwing problems described earlier, and the student might want to think things through in a hypothetical situation. Such an approach might be tutor specific and be used as a strategy to encourage deeper level thinking.
3. The Why2 corpus is *computer mediated* discourse. Because the dialog is mediated by computers, tutors and students need to establish more common ground. Moreover, whereas in face-to-face dialog partners can easily interrupt, in computer mediated discourse hypothetical reasoning may have to be spelled out more.
4. The Why2 corpus is *formal discourse*. Compared to spoken dialog, the Why2 corpus could be considered formal with students not knowing their conversational partner, not being able to interrupt the dialog partner, and being required to communicate in written form. It can thus be argued that in more formal discourse (cf. Finegan 1982) conditionals are more frequently found.

Table 3
Overview of corpora

	Why 2	Santa Barbara	Algebra tutor	CIRCSIM tutor	AutoTutor	JURIS
Setting	Dialog	Dialog	Dialog	Dialog	Dialog	Monolog
Type	Tutoring	Natural	Tutoring	Tutoring	Tutoring	Natural
Interface	Human–human	Human–human	Human–human	Human–human	Human–computer	Human
Formality	'Formal'	Informal	Informal?	'Formal'	Informal	Formal
Domain	Physics	Miscellaneous	Mathematics	Medical	Computer literacy	Legal
Freq words	69,965	66,385	4735	72,971	178,429	670,861
Freq turns	3378	7557	380	6903	18,738	n/a

5. The Why2 corpus is *domain specific*. A high frequency of conditionals is dependent on the domain. The Why2 corpus contains physics jargon, which depends on a more frequent use of conditionals.

Note that these hypotheses are not mutually exclusive. That is, if differences cannot be attributed to dialog, this does not imply that differences cannot be attributed to computer mediated tutoring dialog in a specific domain. However, for reasons of clarity, the analyses reported in Study 2 will treat each hypothesis independently. These five hypotheses were tested by comparing the frequency of conditionals in the Why2 corpus with their frequency in other corpora. An overview of the corpora and their features is given in Table 3.

Hypothesis 1 *(Conditionals are Frequent in Dialog.).* If the higher frequency of conditionals in the Why2 corpus can be attributed to dialog, no difference is predicted to be found between the human–computer Why2 corpus and human–human dialog. A corpus that represents human–human dialog is the Santa Barbara Corpus of Spoken American English (Du Bois, Chafe, Meyer, & Thompson, 2000). It is based on hundreds of recordings of natural speech from various parts of the United States, consisting of conversations, gossips, arguments and on-the-job talks. The corpus is diverse, both in content of the conversation and age, occupation, and class of the participants.

Contrary to the prediction, a significant difference was found ($\chi^2 = 87.68$, $p < 0.001$) between the number of normalized conditionals in the two corpora with fewer conditionals in the Santa Barbara corpus (375) than in the Why2 corpus (679). The same pattern was found for the distribution of conditionals, with 0.9 (raw count: 6) different conditionals in the Santa Barbara corpus, compared to 1.86 (13) in the Why2 corpus. This result appears to falsify the hypothesis that the high frequency of conditionals physics tutoring dialog are more frequent due to the dialog. Though dialog is not a single, uniform category and different types of dialog have different characteristics, the observed frequency at least motivates us to also consider alternative factors.

Hypothesis 2 *(Conditionals are Frequent in Tutoring.).* According to the second hypothesis, more conditionals are to be found in tutoring dialog than in other kinds of dialog, which would explain the high frequency of conditionals in the Why2 corpus compared to both non-tutoring corpora like the BNC and Santa Barbara frequencies. In order to test this hypothesis, we compared the Why2 corpus with the Algebra Tutoring Corpus (Heffernan, 1998). This corpus consists of tutorial dialogs between an experienced math tutor and 8th grade students on algebra symbolization. Tutor and students communicated face-to-face in a classroom. They were videotaped and the speech was transcribed. To write down the symbolizations, the student could use a blank sheet of paper. Each session lasted approximately 1 hour, and consisted of 17 problems. Although language used in algebra is constrained by rules and conventions, the age of the students made the dialog less formal than expected. Table 4 shows that the number of normalized conditionals in this corpus was significantly lower (570; $\chi^2 = 84.42$, $p = 0.03$) than in the Why2 corpus, suggesting that Why2's frequency cannot be attributed to the tutoring setting. The type–token ratio, however, did not reveal any difference (0.42–0.42) possibly due the relatively small size of the algebra corpus. As in the first hypothesis, no justice is done to the diversity of the tutoring corpora and not all variables are held constant; however, the observed results motivate us to consider alternative hypotheses.

Hypothesis 3 *(Conditionals are Frequent in Computer Mediated Discourse.).* The comparison between the frequency of conditionals in the Why2 corpus and the Algebra Corpus already shows that the high frequency cannot be explained by tutoring dialog. To further confirm this conclusion, and to test whether the frequency can be attributed to the

Table 4
Frequencies conditionals corpora

	Why 2	Santa Barbara	Algebra Tutor	CIRCSIM tutor	AutoTutor	JURIS
Freq words	69,965	66,385	4735	72,971	178,429	670,861
Freq turns	3378	7557	380	6903	18,738	n/a
Freq conditionals	475	249	27	366	301	2832
Norm by words	679	375	570	502	169	422
Norm by turns	14.06	3.29	7.11	5.30	1.61	n/a
Types	13	6	4	4	3	17
Type/token	0.42	0.32	0.42	0.23	0.19	0.36

Frequencies are normalized per 10,000 words and 100 turns; type–token ratio is based on the log of types and tokens.

computer mediation, the Why2 corpus was compared with the AutoTutor computer literacy corpus and the CIRCSIM physiology corpus, both computer mediated and – at least for the tutor – scripted tutoring corpora.

AutoTutor is a conversational agent that assists students in actively constructing knowledge by holding a conversation with mixed initiative dialogs. AutoTutor has currently been developed for the domains of computer literacy and physics with the possibilities of development for a range of other domains. The system has a talking head that interacts with the student, a curriculum script that guides the pedagogical strategies of the tutor, and a range of computational linguistic modules that allow interaction with the tutor similar to that with a human (see Graesser et al., 2001). The AutoTutor transcripts used here are student interactions with the computational tutor on the subject of computer literacy. The student can type to AutoTutor, who replies by facial expressions, gestures, speech, and written language. The Why2 transcripts had a significantly larger number of normalized conditionals than the AutoTutor transcripts (169; $\chi^2 = 306.72$, $p < 0.001$). In addition the type–token ratio was lower in the AutoTutor transcripts (0.19) than in the Why2 corpus (0.42).

The second corpus, based on transcripts from the CIRCSIM tutor, is also composed students' interaction with an intelligent tutoring system, but for the domain of cardiovascular physiology. The system carries out a conversational dialog with the user by employing a set of tutoring tactics that mimic those used by two expert human tutors (Evens et al., 2001). Again, the Why2 corpus had a larger number of normalized conditionals ($\chi^2 = 26.52$, $p < 0.001$) than the 502 in the CIRCSIM transcripts. Furthermore, the type–token ratio was larger in the Why2 (0.42) than in the CIRCSIM corpus (0.23).

In consideration of this analysis, which demonstrates that the Why2 corpus has more conditionals than AutoTutor or the CIRCSIM tutoring transcripts, it appears that Why2's high frequency of conditionals may not be explained either by dialog, tutoring or computer mediated conversations.

Hypothesis 4 *(Conditionals are Frequent in Formal Discourse.).* The fourth hypothesis predicts that the higher frequency of conditionals can be explained by the formality of the discourse. It might be the case that in informal discourse fewer conditionals are used than in formal discourse. This would then explain why the Santa Barbara corpus has fewer conditionals than the Why2 corpus, and would possibly explain why the effect cannot be explained by the tutoring environment. For a corpus of formal discourse a random selection of the Justice Department Retrieval and Inquiry System (JURIS) was selected, which contains documents on case law, regulations, international agreements, tax law, etc (Canavan & Morgovsky, 1998). A comparison of the Why2 corpus with the JURIS corpus again showed that the number of normalized conditionals was higher in the Why2 corpus than in the JURIS corpus (422; $\chi^2 = 59.99$, $p < 0.001$). A similar pattern was found for the type–token ratio, with 0.36 for JURIS and 0.42 for Why2. It can therefore be concluded that formal discourse does not explain the higher frequency in Why2.

An overview of the results for the four hypotheses is presented in Table 4.

If dialog, tutoring, computer mediation and formality cannot explain the high frequency in the Why2 corpus, what can? The corpus differs in one important aspect from all other corpora: its domain. This brings us to the fifth hypothesis: that the high number of conditionals can be explained by the domain. This hypothesis was tested next.

Hypothesis 5 *(Conditionals are Frequent in Physics.).* The Why2 corpus consists of questions and answers in the domain of physics. To answer the question whether it is the physics domain that predicts the high frequency, five

Table 5
Overview physics corpora

	Why 2	Einstein	Hewitt	Crowell	BEE	Physics 1
Setting	Dialog	Monolog	Monolog	Monolog	Dialog	Dialog
Type	Tutoring	Academic	Text book	Text book	Tutoring	Tutoring
Interface	Written–written	n/a	n/a	n/a	Written–written	Spoken–spoken
Domain	Conceptual physics	Relativity	Conceptual physics	Conceptual physics	Electricity corpus	Conceptual physics
Freq words	69,965	29,379	47,473	290,358	40,195	114,797
Freq turns	3378	n/a	n/a	n/a	2459	11,496

Frequencies are normalized per 10,000 words and 100 turns; type–token ratio is based on the log of types and tokens.

physics corpora were compared with the Why2 corpus: Einstein's *Relativity: The Special and General Theory* (1920), Hewitt's *Conceptual physics* (1998), Crowell's *Light and Matter* (1998), BEE, a tutoring corpus on Basic Electricity and Electronics (Rosé, DiEugenio, & Moore, 1999), and Physics 1, a human–human tutoring corpus on conceptual physics (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003) (see Table 5).

Einstein's famous text describes Relativity Theory. Hewitt's text is a non-computational textbook on the central concepts of physics organized by six major units: Mechanics, Properties of Matter, Heat, Sound and Light, Electricity and Magnetism, and Atomic and Nuclear Physics. Crowell's text contains six volumes of a comprehensive text book aimed at students in technical or life science majors. In the Basic Electricity and Electronics Corpus, students interacted with a human tutor while working through a computer based basic electricity and electronics curriculum. Lastly, the Physics 1 corpus is a tutoring corpus on conceptual physics wherein tutor and students communicated through a conference phone that only let one person talk at a time. If the higher frequency of conditionals can be explained by the physics domain, we expect not to find differences between the Why2 corpus and these five corpora. Indeed, a Chi squared test revealed no difference between the Why2 corpus and Einstein's study, the Hewitt text, or the two tutoring corpora (all p's > 0.05). The type–token ratio showed similar patterns for the Why2 and Hewitt texts (0.42 and 0.41). Except for the Physics 1 corpus (0.36) the two other corpora used more different conditionals than the Why2 corpus (Einstein: 0.50; BEE: 0.73).

Contrary to what was predicted the Crowell text did have fewer conditionals than Why2 ($\chi^2_{(1)} = 10.63$, $p < 0.001$). Although this would falsify the hypothesis, it needs to be noted here that this corpus is different from all other corpora in the number of illustrative texts, text book questions and answers. In other words, there is a lot of non-physics language present in the textbook. In order to verify this, the corpus was sanitized by removing information non-relevant to the physics domain. The remaining text discussed the core concepts in physics. A comparison of the number of conditionals now showed 11 different conditional types in the sanitized Crowell corpus, with a total normalized frequency of 657. A Chi squared test revealed that this frequency was not different from the Why2 corpus ($p > 0.7$). Furthermore, the type–token ratio was comparable between the Crowell texts (0.41 and 0.47) and Why2 (0.42).

It appears that the high frequency of conditionals in the Why2 corpus can be attributed to the physics domain and its language. Furthermore, whereas in the corpora used to test the first four hypotheses the type–token ratio was smaller than in Why2, the type–token ratio was equal or larger in the physics corpora (Tables 5 and 6).

Table 6
Frequencies conditionals physics corpora

	Why 2	Einstein	Hewitt	Crowell[a]	BEE	Physics
Total words	69,965	29,379	47,473	290,358 (24,371)	40,195	114,797
Total turns	3378	n/a	n/a	n/a	2459	11,496
Freq conditionals	475	173	284	1638 (160)	259	722
Normalized by words	67.9	58.9	59.8	56.4 (65.7)	64.4	62.9
Normalized by turns	14.06	n/a	n/a	n/a	10.53	6.28
Types	13	13	10	19 (11)	11	11
Type/token	0.42	0.50	0.41	0.40 (0.47)	0.43	0.36

Frequencies are normalized per 10,000 words and 100 turns; type–token ratio is based on the log of types and tokens.
[a] Results in brackets are frequencies from the sanitized version of Crowell.

To test this hypothesis further, we analyzed conditionals in the Michigan Corpus of Academic Spoken English (MICASE; Simpson, Briggs, Ovens, & Swales (2002)). This corpus is a spoken language corpus of about 1.7 million words (approximately 200 h) of speech recorded at the University of Michigan. The corpus includes monologic and interactive speech by undergraduate and graduate students, junior and senior faculty and staff, and native and non-native speakers of English. Forty-five disciplines (departments) are represented in the corpus. Examples of these disciplines include anthropology, chemistry, physics, economics, geology, math, and sociology. We selected the 13 most frequent conditionals in the whole MICASE corpus (assume, expect, given, hypothetical, if, imagine, inferred, let, postulated, say, suppose, theoretically, understand), and computed their frequencies per text and normalized their counts by the total number of words in the corpus. When the 152 corpora were rank ordered by total frequency, the four physics corpora occurred in the top 40% of all corpora, with the Graduate Physics Lecture in the top 20% and Introduction to Physics lecture being one of the three corpora with the highest conditionals count (Computer Science Office Hours and Graduate Philosophy Seminar being highest). Not surprisingly, a Chi square test revealed a significant difference between the average conditionals in the Physics (130.95, S.D. = 52.36) versus the non-Physics texts (95.15, S.D. = 38.63) ($\chi^2 = 5.44$, $p = 0.02$).

The testing of the five hypotheses has been reported here as a sequence of eliminations, presuming that there is just one factor rather than an interaction of factors. As pointed out earlier, the five hypotheses are not mutually exclusive. At the same time, however, they provide evidence for the fifth hypothesis that conditionals are more frequent in physics texts. Further research should be conducted to ascertain that an interaction of the factors cannot explain the higher frequency of conditionals in the Why2 corpus.

7. Study 3

So far we have shown that conditionals are more frequent in physics discourse, regardless of the setting (dialog/monolog), the type of discourse (tutoring/textbook), formality (formal, informal), language mode (written/spoken). We concluded that this effect can be attributed to the domain of physics. Given this conclusion, the question can be raised whether advanced physics students are more likely to adopt this physics language compared to less advanced students.

In order to answer this question, we took the student turns of the three physics tutoring corpora used in the previous studies: Physics 1, AutoTutor tutoring in the domain of physics, and Why2. For the last two studies pretest student results were available, allowing for a median split of student ability. Students who scored high on the pretest were classified in the high-ability group, those who scored relatively low in the low-ability group. No pretest information was available for the Why2 corpus. A physics expert therefore divided the students into the two groups based on the student performance during the tutoring session. Table 7 shows the results between these two groups of students in each of the three tutoring corpora.

In all three cases, high ability students used a significantly larger number of conditionals than low ability students. In the Why2 corpus this difference is significant ($\chi^2_{(1)} = 8.71, p = 0.003$) in the Physics 1 and AutoTutor physics corpus the difference is marginally significant ($\chi^2_{(1)} = 3.39, p = 0.065$; $\chi^2_{(1)} = 3.5, p = 0.061$, respectively), but the patterns are consistent with a higher frequency for high ability students compared to low ability student.

Table 7
Frequencies conditionals high versus low ability students

	Why2		Physics 1		AutoTutor physics	
	High ability	Low ability	High ability	Low ability	High ability	Low ability
Words	19,932	13,834	51,414	53,401	22,210	22,043
Turns	761	679	9504	11,703	2662	2779
Conditionals	100	83	102	87	77	66
Norm by words	50.2	60.0	19.8	16.3	34.7	29.9
Norm by turns	13.14	12.22	1.07	0.74	2.89	2.37
Types	7	3	4	2	3	6
Type/token	0.42	0.25	0.3	0.16	0.25	0.43

Frequencies are normalized per 10,000 words and 100 turns; type–token ratio is based on the log of types and tokens.

The question can then of course be raised whether this has to do with the fact that we are dealing with high ability students who have better reasoning or writing skills or with the fact that we are dealing with students who have adopted physics language, which contains a large number of conditionals. To make the comparison of the corpora optimal, student turns in the AutoTutor physics corpus were compared with the student turns in the AutoTutor computer literacy corpus. If the high ability physics students use a higher number of conditionals than the high ability computer literacy students, the language of physics is the explanation. If not, student ability is the answer with more advanced students using a higher number of conditionals.

First, it is important to note that no significant effect was found between the ability levels of students in computer literacy. Furthermore, a significant difference was found between high ability physics students and high ability computer literacy students, with a higher number of conditionals for the physics students ($\chi^2_{(1)} = 21.93, p = 0.001$). We can thus conclude that not only do physics texts have a higher number of conditionals, but students who excel in physics also apply the language of physics.

8. Discussion and conclusion

We have argued that conditionals are important for the field of education to investigate because they can reveal reasoning skills and misconceptions in students even though the precise nature of the misconception may be unknown. Instead of using a precise truth-functional definition of conditionals, we defined conditionals as those relations between two clauses that can be substituted by an *if-then* relation without changing the meaning of that relation. In this study, a total of 106 markers could function as conditionals. By analyzing a series of corpora we investigated how conditionals are used in language in general and in educational language in particular. A very specific corpus, physics tutoring dialog, turned out to have a larger number of conditionals than other corpora. This high frequency could not be explained by the dialogic nature of the text, the tutoring environment or the formality of the text. The only remaining explanation considered was the physics domain and its specific linguistic terminology. When this hypothesis was tested by comparing high ability and low ability students in physics in three similar kinds of tutoring corpora, results showed that high ability students in physics use more conditionals than low ability students, suggesting conformity to the language used in this educational register.

Earlier we outlined the reasons why a corpus linguistic study of conditionals is of interest to an educational community: conditionals shed light on language use in teaching; they can highlight misconceptions in students; they can distinguish between shallow and deep reasoning in students; and they can help in the development of intelligent tutoring systems.

The use of conditionals is an indication of the language user's ability to reason about alternatives, uncertainties, and unrealized contingencies. The results reported here suggest that the problem-solving process in physics is explicit. This is in the line of the findings by Larkin, McDermott, Simon, and Simon (1980) who compared the problem solving behaviors of physics experts (faculty and graduate students) and physics novices (beginning students). They found that experts worked from given information to the goal of the problem, whereas novices started with the goal of the problem and worked "backwards" to the given information. Expert physics uses a knowledge development model, wherein conditionals apparently play an important role.

Student must possess the necessary domain knowledge as well as an understanding of general problem-solving processes and heuristics to be good problem solvers (Maloney, 1994). This realization might lead to rather unexpected tasks for the physics instructor, that which Halliday (1994) termed 'language socialization': by learning the language of physics, students can learn physics. That is to say, by being taught to use conditionals, students are encouraged to make their problem-solving processes explicit. Since we know that shallow knowledge students have difficulties making their problem-solving heuristics explicit, whereas deep knowledge students have mastered this procedure (Maloney, 1994), perhaps additional instructional strategies can be implemented. To the number of problem-solving skills proposed to improve student learning (Cummings, Marx, Thornton, & Kuhl, 1999) we thus propose adding a relatively simple linguistic strategy: physics teachers, human instructors and their computational colleagues, could encourage students to use conditionals in solving their physics problems. This would allow students to be trained in understanding how conditionals are used in physics and how they are productive in establishing deep field knowledge.

Acknowledgements

We would like to thank Eric Mathews for his help in the analysis. This research was supported by the National Science Foundation (IIS 0416128, REC 0106965, ITR 0325428, and REC 0633918). Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

References

Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition*, *75*, 237–266.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics. Investigating language structure and language use*. Cambridge: Cambridge University Press.

Bloom, B. S. (1965). *Taxonomy of educational objectives: The classification of educational goals*. New York: David McKay.

Bransford, J. D., Goldman, S. R., & Vye, N. J. (1991). Making a difference in peoples' abilities to think: Reflections on a decade of work and some hopes for the future. In L. Okagaki & R. J. Sternberg (Eds.), *Directors of development: Influences on the development of children's thinking* (pp. 147–180). Hillsdale, NJ: Erlbaum.

Canavan, A., & Morgovsky, P. (1998). *JURIS* (*Justice Department Retrieval and Inquiry System*). CD-ROM. Philadelphia, PA: LDC.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.

Corbett, A. T. (2001). Cognitive computer tutors: Solving the two-sigma problem. In M. Bauer, J. Vassileva, & P. Gmytrasiewicz (Eds.), *User modeling: Proceedings of the eighth international conference, UM2001* (pp. 137–147). Berlin: Springer.

Corbett, A., Anderson, J., Graesser, A., Koedinger, K., & van Lehn, K. (1999). Third generation computer tutors: Learn from or ignore human tutors? In *Proceedings of the 1999 conference of computer–human interaction* (pp. 85–86). New York: ACM Press.

Crossley, S. A., & Louwerse, M. M. (2007). Multi-dimensional register classification using collocations. *International Journal of Corpus Linguistics*, *12*, 453–478.

Crowell, B. (1998). *Light and matter* (6 vols.). http://www.lightandmatter.com/area1.html (accessed August 1, 2005).

Cummings, K., Marx, J., Thornton, R., & Kuhl, D. (1999). Evaluating innovation in studio physics. *Physics Education Research, Supplement to American Journal of Physics*, *67*, S38–S44.

Dancygier, B. (1999). *Conditionals and prediction: Time, knowledge, and causation in conditional constructions*. Cambridge: Cambridge University Press.

Du Bois, J. W., Chafe, W. L., Meyer, C., & Thompson, S. A. (2000). *Santa Barbara corpus of spoken American English Part-I*. CD-ROM. Philadelphia, PA: LDC.

Einstein, A. (1920). *Relativity: The special and general theory*. London: Methuen.

Evens, M. W., Brandle, S., Chang, R.-C., Freedman, R., Glass, M., Lee, Y., et al. (2001). CIRCSIM-tutor: An intelligent tutoring system using natural language dialogue. In *Twelfth Midwest AI and Cognitive Science conference, MAICS 2001* (pp. 16–23).

Ferguson, G. (2001). If you pop over there: A corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, *20*, 61–82.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*, 104–137.

Graesser, A. C., Person, N. K., & Huber, J. D. (1992). Mechanisms that generate questions. In T. E. Lauer, E. Peacock, & A. C. Graesser (Eds.), *Questions and information systems* (pp. 67–187). Hillsdale, NJ: Erlbaum.

Graesser, A. C., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, *22*, 39–51.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.

Halloun, I., & Hestenes, D. (1985). Common-sense concepts about motion. *American Journal of Physics*, *53*, 1056–1065.

Heffernan, N. T. (1998). *Intelligent tutoring systems have forgotten the tutor: Adding a cognitive model of human tutors*. Computer Science Department, Carnegie Mellon University.

Herdan, G. (1966). *The advanced theory of language as choice and chance*. Berlin: Springer Verlag.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141–158.

Hewitt, P. G. (1998). *Conceptual physics*. Reading, MA: Addison Wesley Longman.

Horsella, M., & Sindermann, G. (1992). Aspects of scientific discourse; Conditional Argumentation. *English for Specific Purposes*, *11*, 129–139.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.

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

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.

Knott, A. (1996). *A data-driven methodology for motivating a set of coherence relations*. Unpublished PhD thesis. University of Edinburgh.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342.

Louwerse, M. M. (2002). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, *12*, 291–315.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the twenty-sixth annual conference of the cognitive science society* (pp. 843–848). Mahwah, NJ: Erlbaum.

Louwerse, M. M., & Mitchell, H. H. (2003). Towards a taxonomy of a set of discourse markers in dialog: A theoretical and computational linguistic account. *Discourse Processes*, *35*, 199–239.

Maloney, D. P. (1994). Research on problem solving: Physics. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 327–356). New York, NY: Macmillan.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Mead, R., & Henderson, W. (1983). Conditional form and meaning in economics text. *The ESP Journal*, *2*, 139–160.

Otero, J., Leon, J. A., & Graesser, A. C. (Eds.). (2002). *The psychology of science text comprehension*. Mahwah, NJ: Erlbaum.

Rosé, C. P., DiEugenio, B., & Moore, J. (1999). A dialogue-based tutoring system for basic electricity and electronics. In S. P. Lajoie & M. Vivet (Eds.), *Artificial intelligence in education* (pp. 759–761). Amsterdam: IOS.

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan corpus of academic spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Strawson, P. F. (1986). If and "$\supset$". In R. E. Grandy & R. Warner (Eds.), *Philosophical grounds of rationality. Intentions, categories, ends* (pp. 363–380). Oxford: Clarendon Press.

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, *21*, 209–249.