

The effect of lexical, pragmatic, and morphological violations on reading time and deviance ratings of English and German sentences

AARON L. WICHMAN

Ohio State University, Columbus, Ohio

and

BRIAN M. FRIEL and RICHARD JACKSON HARRIS

Kansas State University, Manhattan, Kansas

Effects on reading time and deviance ratings of word-choice violations were studied at the surface code, propositional textbase, and situation model levels of representation in German and English sentences. Lexical violations (propositional textbase level) such as *The housewife massaged the bread dough* were rated as more deviant from normal language than were pragmatic violations (situation model level) such as *The police officer shot the parking violator*, although the pragmatic violations took longer to comprehend. The addition of a morphological (surface code) violation through the wrong form of an article (e.g., *a ugly car*) decreased reading time but only in cases where that material was deeply processed and where the morphology carried substantial syntactic information (e.g., in German, the definite article carries case, gender, and number information). The results thus confirmed the operation of comprehension strategies at multiple levels of representation and identified some cross-linguistic generality in comprehension and some language differences.

Five levels of representation involved in discourse comprehension have been identified (Graesser, Millis, & Zwaan, 1997). The *surface code* is the actual sequence of words. The *propositional textbase* is the construction of the basic meaning in propositional form. The retrieval of the appropriate knowledge structures with which to interpret that textbase and connect it to known information already in memory involves the *referential situation model*, which comprises the knowledge structures necessary for one to interpret the discourse (Zwaan, 1999; Zwaan & Radvansky, 1998). The *pragmatic communication level* involves the interpretation of the use, or possibly the violation of, pragmatic constraints. For example, an author might describe events in a backward time sequence instead of the usual forward fashion. Lastly, the *text genre* level involves consideration of the type of text of the discourse (e.g., fiction vs. exposition). Processing at any one of these five levels can be thwarted during on-line comprehension if one encounters an unexpected word. The effects of such “surprise words” at the first three of these levels were examined in the present study.

On-line sentence comprehension involves an attempt to integrate the information from a given sentence into a coherent whole as rapidly and as simply as possible. This processing occurs at the multiple levels of representation discussed above. The surface code is involved, for example, at the level of morphological agreement. In English, the indefinite article *a* becomes *an* before a vowel sound. In many languages, an article, adjective, or noun changes through inflection, depending on the gender, case, and/or number of the following noun.

The propositional textbase level is involved, for example, in the semantic conventions regarding the lexical concordance of constituents, which require some co-occurrence restrictions on the choice of verb and object. A student *does* homework, rather than *makes* homework. *We mow* the grass, we don't *shave* the grass. Although a sentence which violated such a lexical convention would often be comprehensible, at least with a bit of effort, it clearly diverges from what is considered “correct” language usage and would be expected to add to processing difficulty.

Other word-choice violations may go against conventions at the situation model level. The propositional content of a sentence can be possible but not highly probable, given our constructed situation model. For example, *The prime minister inaugurated the public toilet* may be seen as less probable than *The prime minister cut the ribbon for the new subway system* because of knowledge about the world—specifically, about the sorts of things prime ministers do. The violation at this situation model

The cooperation of S. Krieger and S. Lis of Justus Liebig University on Experiment 1 is gratefully acknowledged. The comments of Mike Rinck and Tuan Tran on previous drafts of the manuscript are greatly appreciated. Correspondence may be addressed to A. L. Wichman, Department of Psychology, Ohio State University, Columbus, OH 43210, or R. J. Harris or B. M. Friel, Department of Psychology, Kansas State University, 492 Bluemont Hall, 1100 Mid-Campus Drive, Manhattan, KS 66506-5302 (e-mail: wichman.3@osu.edu, rjharris@ksu.edu, bfriel@ksu.edu).

level requires reference to knowledge of the world and thus poses a different sort of comprehension difficulty than the purely lexical concordance case, which may be resolved at the propositional textbase level.

There is evidence for comprehension's being influenced by these various levels of representation. For example, Mitchell (1987) showed the importance of propositional textbase processing, where an optionally transitive verb (a lexical feature) affected reading time for "garden-path" sentences. *After the child had visited the doctor prescribed a course of injections* took longer to read than *After the child had sneezed the doctor prescribed a course of injections*, presumably because participants expected *the doctor* to be the direct object of the optionally transitive verb *visited*, rather than the subject of the next clause. With no comma after *visited*, the lexical information contained in the verb caused an error in parsing. In the other sentence, there was no such problem, because *sneezed* is intransitive and could not claim *doctor* as its direct object.

On the other hand, comprehension is also influenced by what is most plausible in terms of situation models and world knowledge. Crain and Steedman (1985) showed that the sentence *The children taught by Berlitz method passed the test* was judged as more grammatical than *The teachers taught by Berlitz method passed the test*, presumably because of the conventional wisdom that children are more likely than teachers to be taught.

In the present experiments, we examined the effects of three levels of representational violations of word choice on reading times and ratings of grammatical deviance. Specifically, our purpose was to investigate the interaction of lexical (propositional textbase level) and pragmatic (situation model level) violations with morphological violations involving inflected articles in German and English (surface code level). In German, an incorrect declension of the accusative (direct object) definite article was used as the morphological violation, whereas in English, the incorrect indefinite article was used. Two languages (English and German) were studied for purposes of replication beyond one language and also to examine a slightly different sort of morphological violation for the article manipulation. Owing to the nature of the materials, the surface code manipulation (article violation) was varied orthogonally across the lexical and pragmatic violations, allowing an examination of each type of violation by itself but also of interactions of the morphological violation with each of the other two levels.

The importance of studying more than one language is quite clear. In a review of studies and models of parsing, Cuetos, Mitchell, and Corley (1996) have argued that a major issue in developing and testing such models is the degree of universality of the theory. On the one hand, some have argued or assumed that parsing strategies are basically universal across all languages (Frazier, 1987; Inoue & Fodor, 1995; Pritchett, 1992). In these models, language differences are handled essentially by considering such information as data files made available to a

universal parser. Other models argue for various degrees of customization that vary in the extent to which they interpret language differences as "file exchanges" replacing one grammar with another (Bates & MacWhinney, 1987; Clahsen, 1999; Cuetos & Mitchell, 1988; Gilboy & Sopena, 1996; Mazuka & Lust, 1990; St. John & McClelland, 1990). One can never competitively support one type of model over another without examining multiple languages.

Hypotheses

We predicted that lexical and pragmatic violations would have different effects on ratings of deviance from normal language use than morphological violations would. We predicted that lexical violations would be perceived as the most deviant from normal language usage, because they represent an entirely linguistic impropriety. Pragmatic violations, on the other hand, are inconsistent with the world at the level of the situation model rather than with linguistic conventions; thus they should be rated less deviant than lexically violated sentences but more deviant than the control "correct" sentences. The addition of the morphological violation of an incorrect article should increase the deviance ratings across all sentence types, because of its high saliency and very low frequency of occurrence in normal language usage.

In terms of reading times, we predicted that pragmatically violated sentences would be viewed longer than lexically violated sentences before rating, because of the necessity of checking against the real-world knowledge in one's situation model. The lexical violation involves a purely linguistic inconsistency with the internal rules of what lexical items may or may not co-occur. Thus it should take longer to use the situation model to evaluate the real-world probability of a sentence (pragmatic violation) than to read and evaluate a sentence based on verb-object disagreement (lexical violation). For this reason, although we predicted that pragmatically violated sentences would be rated as less deviant from normal than lexically violated sentences, we also predicted that they would take longer to read, yielding the longest predicted comprehension viewing times of the three basic sentence types. We hypothesized that lexically violated sentences take less time to read than pragmatically violated ones, because only propositional textbase information had to be accessed to judge these sentences. Control (violation-free) sentences were predicted to be most quickly viewed and evaluated.

The morphological violation manipulation was necessarily different for the German (Experiment 1) and English language cases (Experiments 2 and 3). For the German morphological violation, the article in front of the direct object of the transitive verb appeared in the nominative, rather than the correct accusative, case. For the English violation, the incorrect version of the indefinite articles *a* or *an* was used. This manipulation was chosen to be the closest English surface code analogue to the case violation in German. It was hypothesized for the German

language study that the incorrect declension would increase reading time. This would occur because, although the lexical information in the verb was transitive, leading the participant to expect a direct object, the article—the German equivalent of *the*—nevertheless signaled a new subject (it was in the nominative, not accusative, case), not an object of the just-read transitive verb. For the English language study, it was predicted that incorrect indefinite article manipulation would have a similar, although somewhat weaker, effect. This is because the usage of *a* or *an* does not signal the grammatical case of the following noun, but rather only whether that noun begins with a consonant or a vowel. It thus carries less information and requires less extensive processing.

The effect of morphological article (surface code) violation on reading time across the three different sentence types was predicted to be constant, adding a comparable parsing difficulty to the evaluation of the various sentence types.

EXPERIMENT 1

Method

Participants. The participants were 20 native German-speaking right-handed volunteers from 20 to 28 years old, 10 men and 10 women. All had successfully completed at least the German college preparatory school, or Gymnasium. Sixteen participants had had at least 1 year of University education. Data were collected in Giessen, Germany.

Materials. The 120 stimuli were six types of sentences, 20 of each type. All had the following basic form: subject + finite transitive verb in the present tense + direct object compound noun. The verb described an activity typical for the subject, with a goal of making all sentences conceptually simple to process. Subjects were always singular and named people designated as professionals or in family roles: criminal, collector, hunter, barber, or pilot; son, mother, or grandfather; and so forth. For example, one sentence was *Der Apotheker empfiehlt den Hustensaft* [The pharmacist recommends the cough syrup], where *Der Apotheker* [The pharmacist] is the subject, *empfiehlt* [recommends] the correctly conjugated verb, and *den Hustensaft* [cough syrup] the direct object, with the masculine singular definite article *den* in the accusative case.

This basic form was also used for the other two types of sentences. One type altered the *lexical consistency* of the verb–object pair, representing a violation at the propositional textbase level of representation. For example, in the sentence *Die Hausfrau massiert den Kuchenteig* [The housewife massages the cake dough], *massiert* [massages] is not normally paired with the object *Kuchenteig* [cake dough]. Although the intended semantic content of the sentence is at least calculable, the lexical items used to express this content are paired in a fashion that is not quite correct.

In contrast, *pragmatically violated* sentences expressed content that was highly improbable for anyone drawing upon a situation model of knowledge of the world but which did not violate word usage rules. For example, consider *Der Polizist erschießt den Falschparker* [The policeman fatally shoots the parking violator]. Although a policeman could conceivably shoot someone fatally, shooting a parking violator would be highly unlikely, given what most people believe about the way police in the “real world” work. Another example of a pragmatic violation is *Der Bettler heiratet die Kinderärztin* [The beggar marries the pediatrician]. The beggar could marry anyone, but not very likely a pediatrician!

In addition to these three forms, we constructed an equal number of additional sentences in each of the three basic forms (morpho-

logical violations) by placing an incorrectly declined definite article in front of the direct object of the verb.

In German, there are several possible definite articles, in contrast to only *the* in English. German articles must match their nouns in gender (masculine, feminine, or neuter), number (singular or plural), and case (nominative, accusative, dative, or genitive). For the present study, all of the morphological violations used the masculine nominative (subject) singular form *der* in front of a masculine singular accusative (direct object) noun, which should correctly take the accusative article *den*. The use of the masculine form of the article most clearly drew attention to this violation, because only this gender has distinct nominative and accusative definite articles.

In summary, six types of sentences were used:

1. Control: *Der Apotheker empfiehlt den Hustensaft* [The pharmacist recommends the cough syrup].
2. Control + morphological violation: *Der Apotheker empfiehlt der Hustensaft*.
3. Lexical violation: *Die Hausfrau massiert den Kuchenteig* [The housewife massages the cake dough].
4. Lexical violation + morphological violation: *Die Hausfrau massiert der Kuchenteig*.
5. Pragmatic violation: *Der Polizist erschießt den Falschparker* [The policeman fatally shoots the parking violator].
6. Pragmatic violation + morphological violation: *Der Polizist erschießt der Falschparker*.

Although for illustration above the same sentences are shown with and without the morphological violation, any given participant saw only one of the two sentences of each pair.

Procedure. Participants were tested individually. Although the experimenter was nearby, participants were told to “pretend you are alone and have to complete this program all by yourself.” This instruction was used to discourage questions and conversation with the experimenter, which could greatly inflate reading times. Consequently, only 2 out of 1,200 individual reaction times had to be discarded because of questions by participants.

A computer program presented half of the 120 sentences to each participant, one of two previously randomized-within-constraints orderings of the material (10 blocks \times 6 sentence types = 60 sentences). Each sentence type appeared once in its block of 6 sentences, and lists were constrained in such a way that no two sentences of the same type followed one another directly in the presentation list. Sentences were presented in bold 14-point type on a laptop computer.

The participant was told that “a list of more or less incorrect sentences” would be shown, randomly divided into three equal-sized subgroups. Participants were instructed to rate the sentences in terms of their “deviation from normal language usage,” on a Likert scale from 1 to 9, with 1 representing *no* deviation from normal language use and 9 representing a *strong* deviation from normal use. Participants were told to “Use the criteria that seem to make the most sense to you” for assessing deviance.

Participants were instructed to call up each sentence by pressing the space bar, whereupon the sentence appeared on the screen. They were to read the sentence and decide on a deviance rating. Once they had done this, they were to release the space bar, whereupon the sentence disappeared from the screen and they entered their rating (1–9) via the number keys along the top row of the keyboard. Participants used the index finger for all responding. To view the next sentence, the space bar was pressed, and the process was repeated. This continued until all 60 sentences had been viewed and rated. During both the reading of the sentence and the entering of the rating, the rating scale appeared at the bottom of the screen, with the end points marked and its numbers from 1 to 9 roughly in line above those numbers on the keyboard. Although the sentence disappeared upon release of the space bar, the rating scale itself re-

mained on the screen. In order to minimize the startling onset of the sentences, the program presented the sentences 0.25 sec after the participant had depressed the space bar. Practice sentences were given.

Two dependent measures were recorded. The participant's deviance rating was noted, along with the reading time (interval measured from the moment the sentence appeared until the space bar was released) in milliseconds for each trial.

Results

Trimming of reading time data. To minimize the influence of outliers, 10% of the most extreme values from the reading time measure were trimmed. This trimming procedure was carried out separately for each sentence type. Of the 1,200 reading times recorded from the 20 participants (20 per sentence type), the top and bottom 5% were trimmed and replaced with the mean for each participant's trimmed distribution for that particular sentence type. Out of a total of 1,200 data points, 120 were thus removed and replaced with the mean of their trimmed distributions. A chi square analysis on the frequencies of trimmed readings [$\chi^2(2, N = 120) = 5.48, p > .05$] across the six conditions revealed no significant differences between the number of values trimmed as a function of sentence type or article condition.

Reading times. Table 1A displays the mean reading times for each sentence and article condition. For participants (F_1) and items (F_2), 2 (article) \times 3 (sentence type) repeated measures analyses of variance (ANOVAs) were conducted. A main effect of article was obtained in both analyses [$F_1(1,19) = 10.08, p < .005, \eta^2 = 0.347$; $F_2(1,9) = 81.26, p < .0001, \eta^2 = 0.900$], as morphologically violated sentences yielded shorter reading times ($M = 5,140.24, SD = 1,646.31$) than did sentences with the correct article ($M = 6,089.54, SD = 2,033.09$). A significant main effect of sentence type was also obtained in both analyses [$F_1(2,38) = 11.59, p < .0001, \eta^2 = 0.379$; $F_2(2,18) = 9.98, p = .0012, \eta^2 = 0.526$]. Simple contrasts were conducted on the means for sentence type. Pragmatically violated sentences took significantly longer to read ($M = 6,260.46$ msec, $SD = 2,074.94$) than did either control sentences [$M = 5,248.16$ msec,

$SD = 1,892.43$; $F(1,19) = 15.69, p < .001, \eta^2 = 0.452$] or lexically violated sentences [$M = 5,336.56$ msec, $SD = 1,585.79$; $F(1,19) = 16.27, p < .001, \eta^2 = 0.461$], which did not differ from each other. There was no sentence type \times article interaction.

Deviance ratings. Two 2 (article) \times 3 (sentence type) repeated measures ANOVAs (for participants and for items) tested the effects of article and sentence type on deviance rating. Results appear in Table 1B. The article \times sentence type interaction was significant in both analyses [$F_1(2,38) = 29.01, p < .0001, \eta^2 = 0.604$; $F_2(2,18) = 20.00, p < .0001, \eta^2 = 0.690$]. A series of nine dependent t tests was performed, with a Bonferroni adjustment of the alpha level ($.05/9 = .0056$), to test pairwise differences within the levels of each factor. All comparisons were significant except for the comparison of the ratings for lexically and pragmatically violated sentences with article violations.

Significant main effects of article were also obtained in both analyses [$F_1(1,19) = 55.99, p < .0001, \eta^2 = 0.747$; $F_2(1,9) = 232.16, p < .0001, \eta^2 = 0.963$]. Sentences with article violations received higher deviance ratings ($M = 6.90, SD = 2.24$) than did sentences with no such violations ($M = 3.36, SD = 2.04$). The main effect of sentence type was also significant in both analyses [$F_1(2,38) = 66.18, p < .0001, \eta^2 = 0.777$; $F_2(2,18) = 92.93, p < .0001, \eta^2 = 0.912$]. Simple contrasts were conducted on sentence type means (control, $M = 3.28, SD = 2.69$; lexically violated, $M = 5.95, SD = 1.81$; pragmatically violated, $M = 4.96, SD = 2.33$). All pairwise comparisons were significant [control vs. lexically violated; $F(1,19) = 364.68, p < .0001, \eta^2 = 0.950$; control vs. pragmatically violated, $F(1,19) = 33.11, p < .0001, \eta^2 = 0.635$; lexically violated vs. pragmatically violated, $F(1,19) = 16.184, p < .001, \eta^2 = 0.460$].

Discussion

As predicted, the main effect of sentence type on deviance rating was obtained. Lexically violated sentences were rated as more deviant than either the control or the

Table 1A
Mean Reading Times (in Milliseconds, With Standard Deviations): Experiment 1

Article Condition	Sentence Type						<i>M</i>	
	Control		Lexical Violation		Pragmatic Violation			
	RT	<i>SD</i>	RT	<i>SD</i>	RT	<i>SD</i>	RT	<i>SD</i>
Correct	5,615	1,749	5,720	1,717	6,934	2,387	6,090	2,033
Violated	4,880	2,002	4,953	1,380	5,587	1,479	5,140	1,646
<i>M</i>	5,248	1,892	5,337	1,586	6,260	2,075		

Table 1B
Mean Deviance Ratings (With Standard Deviations): Experiment 1

Article Condition	Sentence Type						<i>M</i>	
	Control		Lexical Violation		Pragmatic Violation			
	DR	<i>SD</i>	DR	<i>SD</i>	DR	<i>SD</i>	DR	<i>SD</i>
Correct	1.27	0.28	5.29	1.45	3.53	1.50	3.36	2.04
Violated	5.28	2.52	6.61	1.92	6.39	2.13	6.90	2.24
<i>M</i>	3.28	2.69	5.95	1.81	4.96	2.33		

Note—For deviance ratings, 1 = no deviation from normal, 9 = strong deviation from normal.

pragmatically violated sentences. This suggests that it was divergence from normalcy in the propositional textbase more than the situation model level that was taken into account when participants were choosing a deviance rating. When a morphological article violation was added, it increased each sentence type's deviance rating. With the control sentences, this was not surprising; the article in the modified condition was the only thing wrong with them. However, in the other two conditions, the incorrect article added to the sentences' baseline (*no violation*) deviance rating.

The predictions regarding reading times for the three sentence types were also generally confirmed. Pragmatically violated sentences took longer to read than either the lexically violated or control sentences, which did not differ from each other. The necessity of referring to the situation model level to evaluate the pragmatically violated sentences required more time. Contrary to prediction, the control sentences took no less time to read and evaluate than did the lexically violated ones. Apparently the controls were very quickly judged as nondeviant and the lexical violations quickly judged as highly deviant, in neither case requiring much, if any, reference to a situation model.

Contrary to expectations, the prediction that the article violation would increase reading time was not confirmed. In fact, the opposite was true: The violation had a constant effect across the three sentence types of *reducing* the time required to decide on a deviance rating (see Tables 1A–1B). Although it may at first appear surprising that the addition of an article violation should actually *reduce* the amount of time required for one to process a sentence, the incorrectly declined article was probably used as a decision-making cue by participants to determine a deviance rating. Participants saw the incorrect article and quickly assigned more extreme deviance ratings to morphologically violated sentence types than to their non-morphologically violated counterparts. Once participants saw the article error, they required relatively little additional processing in order to make the deviance judgment. In the sentences without the morphological violation, however, meaningful processing had to continue, moving at least to the propositional textbase level for the lexical violations, and, in the case of the pragmatic violations, to the situation model level of representation.

It is quite likely that a significant amount of parallel processing of the different levels occurred. The surface code information, including issues such as article-noun agreement, might be processed in parallel with propositional textbase information. Processing making reference to the situation model might or might not begin in parallel with the other levels but was likely to continue after processing at the surface code and propositional textbase levels were complete. In the present experiment, encountering the accusative article violation might have tipped the decision processes heavily toward a highly deviant rating. In the case of the pragmatic violation plus

the article violation, however, processing must have continued with reference to the situation model, given that there was probably some sense of "incorrectness." This was because the pragmatic violations referenced unlikely events. To determine whether the events were actually plausible, processing had to continue. At the level of propositional textbase representation, that sense of "incorrectness" brought on by pragmatic violations was less than that obtained with the lexical violations. Thus the longest time was required for the pragmatic violations.

EXPERIMENT 2

Although the results from Experiment 1 were robust, their generalizability was still unknown. The effect of the lexical and pragmatic violations could be limited to German, and the morphological violation effect could be limited to this particular accusative case definite article violation. For these reasons, Experiment 1 was conceptually replicated in English. The lexical and pragmatic violations would translate clearly, but the accusative case violation was of course not replicable in English, because there are no article case markers in English. In an attempt to find a form as morphologically comparable as possible, the indefinite article was selected. In English, the singular indefinite article *a* changes to the form *an* before a succeeding vowel. Unlike the German case declension, however, this change is purely for sound quality and carries no information about case or gender. Thus, although this surface morphological manipulation was very similar to the *der/den* change in German, it was a linguistically shallower change, since it carried only surface phonological, and no syntactic, information. Thus this manipulation is structurally identical on a superficial level to that used in Experiment 1, but it did not carry as much information. Thus, we predicted that the effects obtained in Experiment 1 would replicate, though in a weaker form, for the morphological violation variable.

Method

Participants. The participants were 20 native English-speaking Americans from 17 to 28 years old, 10 men and 10 women. They were college psychology students, and they received partial course credit for their participation.

Materials. This study essentially replicated Experiment 1 in English, although the sentences were not always literal translations of those in German. Control sentences, lexical violations, and pragmatic violations were used, each with a second condition of morphologically violated article usage. Instead of the German definite article *der* being used incorrectly, the singular indefinite article *a* or *an* was used inappropriately. Sentences took the form of singular subject + transitive verb + *a* or *an* + adjective + direct object—for example, *The son loses a school bag*, *The barber trims an ugly goat*. In order to achieve the desired morphological violation most comparable to the German nominative-accusative definite article case violation, *a* and *an* were used incorrectly—*The son loses a school bag*, *The barber trims a ugly goat*. There were thus the same six types of sentences used in Experiment 1:

1. Control sentences: *The mechanic checks a fuel line.*
2. Control sentences with morphological violation: *The mechanic checks an fuel line.*
3. Lexically violated sentences: *The mother launders a crying baby.*
4. Lexically violated sentences with morphological violation: *The mother launders an crying baby.*
5. Pragmatically violated sentences: *The nun pays an undulating stripper.*
6. Pragmatically violated sentences with morphological violation: *The nun pays a undulating stripper.*

Procedure. The participants completed the experimental task individually under the same conditions as in Experiment 1. They were told that they would be "... asked to read and to judge different sentences that are shown on a computer screen. The results will help increase understanding of human language processing." Following this, additional procedural instructions were given, and five practice sentences of varying degrees of violation were presented. As in Experiment 1, the length of time that it took the participants to view each sentence before releasing the space bar and entering the rating (reading time) was recorded, as well as their deviance ratings.

Counterbalancing was used exactly as in Experiment 1. The 60 sentences contained 20 of each of the three basic sentence types, with half of each 20 using *a* or *an* incorrectly and the other half using these indefinite articles correctly. Each participant saw all 60 sentences, with one half of the sentences in each of the three sentence types containing an incorrect article. As in Experiment 1, a very small number of reading times could not be used, owing to participant questions or some distraction in the environment. For Experiment 2, as well as the following Experiment 3, the number of these values was less than 1% of 1,200 total data points.

Results

Reaction times were trimmed in the same manner as in Experiment 1. A chi-square analysis of the frequencies of the trimmed reading times [$\chi^2(2, N = 120) = 1.07, p > .05$] across the six conditions again revealed no significant differences between the numbers of values trimmed as a function of sentence type or article condition.

Two 2 (article) \times 3 (sentence type) repeated measures ANOVAs, for participants (F_1) and for items (F_2), were conducted for each dependent variable (reading time and deviance rating). Reading times are presented in Table 2A. Unlike in Experiment 1, neither article nor sentence type nor their interaction was significant in either the participants or the items analysis.

For deviance ratings, the main effect of article was significant in both analyses [$F_1(1,19) = 10.29, p < .005, \eta^2 = 0.351; F_2(1,9) = 112.78, p < .0001, \eta^2 = 0.926$]. Sentences with article violations were given higher deviance ratings ($M = 5.36, SD = 2.15$) than were those with the correct article ($M = 3.75, SD = 2.17$). Sentence type was also significant in both analyses [$F_1(2,38) = 30.98, p < .0001, \eta^2 = 0.620; F_2(2,18) = 64.33, p < .0001, \eta^2 = 0.877$]. Simple contrasts were conducted on the sentence type means collapsed over article (control, $M = 3.04, SD = 2.10$; lexically violated, $M = 5.07, SD = 1.91$; pragmatically violated, $M = 5.55, SD = 2.09$). Control sentences were rated as significantly less deviant than either lexically violated [$F(1,19) = 46.68, p < .0001, \eta^2 = 0.719$] or pragmatically violated [$F(1,19) = 35.65, p < .0001, \eta^2 = 0.652$] sentences, which did not differ from each other.

Also, as in Experiment 1, the article \times sentence type interaction was significant in both analyses [$F_1(2,38) = 6.02, p < .01, \eta^2 = 0.241; F_2(2,18) = 11.71, p < .001, \eta^2 = 0.565$]. Nine dependent *t* tests were performed, with a Bonferroni alpha adjustment ($.05/9 = .0056$), to test pairwise differences among cell means within the levels of each factor. Within the article violation condition, control sentences were rated as significantly less deviant than lexically violated or pragmatically violated sentences, which did not differ from each other. The same pattern was found for deviance ratings of sentences without article violations. For the control condition, sentences with article violations were rated as significantly more deviant than were those with the correct article. No other comparison was significant. The results appear in Table 2B.

Table 2A
Mean Reading Times (in Milliseconds, With Standard Deviations): Experiment 2

Article Condition	Sentence Type							
	Control		Lexical Violation		Pragmatic Violation		M	
	RT	SD	RT	SD	RT	SD	RT	SD
Correct	5,776	1,599	5,529	1,492	5,876	1,533	5,727	1,523
Violated	5,189	1,562	5,805	2,123	5,629	1,782	5,541	1,824
M	5,483	1,588	5,667	1,817	5,753	1,646		

Table 2B
Mean Deviance Ratings (With Standard Deviations): Experiment 2

Article Condition	Sentence Type							
	Control		Lexical Violation		Pragmatic Violation		M	
	DR	SD	DR	SD	DR	SD	DR	SD
Correct	1.86	1.27	4.46	1.63	4.92	2.14	3.75	2.17
Violated	4.22	2.13	5.68	2.02	6.18	1.90	5.36	2.15
M	3.04	2.10	5.07	1.91	5.55	2.09		

Note—For deviance ratings, 1 = no deviation from normal, 9 = strong deviation from normal.

Discussion

Although the deviance rating data largely replicated those obtained in the first experiment, the reading time data did not. For the pragmatic violation and control conditions, the sentence was read and the deviance rating was made faster in the article violation condition, although the difference was not significant as it had been in Experiment 1.

Before further interpreting these results, we decided to ensure that participants were truly processing the sentences fully semantically. The lack of differences in reading times across sentence type might have been explained by a general failure to take the task seriously and fully process the sentences. Unlike with the German materials in Experiment 1, where presumably more time was required for one to parse and comprehend the sentence to a sufficient degree to notice the incorrect declension, the English indefinite article violation in Experiment 2 could have been identified by one's noticing only that the article and its following noun did not go together. When used in the nominative case rather than in the accusative (the latter was the case in Experiment 1), the article–noun combination in the article violation condition for German sentences (e.g., *der Hustensaft*) is a correct combination, whereas that for the English sentences (e.g., *an school bag*) can never be considered correct. Thus, the English sentences in Experiment 2 might have been processed only superficially, enough for one to catch the morphological violation but not enough to attend to the lexical or pragmatic violations, whereas the German participants in Experiment 1 were required to process the sentences more deeply in order to catch the article violation.

However, the deviance ratings in Table 2B, suggest that the participants did in fact notice the lexical and pragmatic violations as well, so that the processing must not have been completely limited to the surface code level. Still, if there had been even some short-circuiting of the propositional textbase or situational model processing relative to Experiment 1, this could explain the failure to replicate the pragmatic-lexical violation difference in reading times obtained in Experiment 1. Thus, in retrospect, Experiment 2 might not have been the fairest test of the predictions. Unfortunately, the reading times for the lexically violated and control conditions obtained in Experiment 2 were greater than those obtained in Experiment 1, providing some evidence against this explanation. However, materials from Experiment 2 were not matched in terms of word length with those in Experiment 1, a fact that could account for the longer reading times in Experiment 2.

An additional possibility was that motivational differences between participant samples might have been contributing to this failure to replicate. In Experiment 2, participants came from an introductory psychology subject pool and were required to participate as part of a course requirement. The German participants in Experiment 1 were asked individually by the experimenter to participate

in a study collecting baseline data for a broader study of mental illness and possible treatments. They participated entirely voluntarily and should have been more motivated and accountable, with the possible consequence that they processed the experimental stimuli more carefully. To explore these questions further, Experiment 3 was conducted, using the same materials as Experiment 2 along with a procedural change to ensure deeper semantic processing, in an attempt to replicate the pattern of results obtained in Experiment 1.

EXPERIMENT 3

In Experiment 3, English-speaking participants were again asked to read and rate each sentence, but after making the deviance rating they were instructed to say what they thought might logically follow the sentence. The purpose was to discourage them from basing the deviance rating largely on the very superficial cues of article–noun sound disharmony without processing the entire sentence and being forced to notice the lexical or pragmatic violation. It was hoped that this manipulation would help ensure that the English sentences with violations would have to be fully processed semantically, more as had naturally occurred with the German sentences in Experiment 1.

Method

In Experiment 3, we tested 20 new native English-speaking participants from 18 to 46 years old, who were from the same pool as in Experiment 2. The sentences and the counterbalancing of the lists were identical to those in Experiment 2.

All aspects of the procedure for Experiment 3 were identical to those of Experiments 1 and 2, with the exception that participants were told,

Once you have entered your [deviance] judgment, tell what might logically follow the sentence you just viewed. For instance, after the sentence: “*The attendant helps an exhausted guest,*” you might say: “*Then the attendant puts the guest in bed for a good night’s sleep*” or “*Then the guest goes to bed.*”

The participants spoke this consequence into a cassette recorder. The experimenter explained that the participant’s comments on the tape would be scored for accuracy, although that never occurred. As in the previous experiments, reading time (in milliseconds) and deviance rating measures were recorded.

Results

The reading time data were trimmed as in Experiments 1 and 2. The chi-square analysis on the frequencies of trimmed reading times [$\chi^2(2, N = 120) = 2.30, p > .05$] across the six conditions once again revealed no significant differences among the number of values trimmed as a function of sentence type or article condition.

Two 2 (article) \times 3 (sentence type) repeated measures ANOVAs were performed on deviance ratings and reading times. As was the case in Experiment 2, there was no significant main effect of article on reading times in either analysis, although there was a main effect of sentence type [$F_1(2,38) = 6.68, p < .005, \eta^2 = 0.260; F_2(2,18) =$

Table 3A
Mean Reading Times (in Milliseconds, With Standard Deviations): Experiment 3

Article Condition	Sentence Type							
	Control		Lexical Violation		Pragmatic Violation		<i>M</i>	
	RT	<i>SD</i>	RT	<i>SD</i>	RT	<i>SD</i>	RT	<i>SD</i>
Correct	11,410	4,236	10,704	4,418	11,943	4,192	11,019	4,263
Violated	10,172	3,759	11,310	3,796	11,649	4,144	11,044	3,889
<i>M</i>	10,291	3,955	11,007	4,077	11,796	4,117		

Table 3B
Mean Deviance Ratings (With Standard Deviations): Experiment 3

Article Condition	Sentence Type							
	Control		Lexical Violation		Pragmatic Violation		<i>M</i>	
	DR	<i>SD</i>	DR	<i>SD</i>	DR	<i>SD</i>	DR	<i>SD</i>
Correct	2.06	1.02	4.52	1.45	5.03	1.82	3.87	1.94
Violated	3.39	1.51	4.96	1.70	5.57	1.92	4.64	1.93
<i>M</i>	2.73	1.44	4.74	1.58	5.30	1.86		

Note—For deviance ratings, 1 = no deviation from normal, 9 = strong deviation from normal.

14.12, $p < .0005$, $\eta^2 = 0.611$]. Data appear in Table 3A. Simple contrasts on the sentence type means collapsed over article (control, $M = 10,290.85$, $SD = 3,954.62$; lexically violated, $M = 11,006.71$, $SD = 4,077.26$; pragmatically violated, $M = 11,796.43$, $SD = 4,116.68$) revealed that pragmatically violated sentences differed from control sentences [$F(1,19) = 11.53$, $p < .005$, $\eta^2 = 0.378$] and from lexically violated sentences [$F(1,19) = 8.65$, $p < .01$, $\eta^2 = 0.313$], which did not differ from each other. The article \times sentence type interaction failed to reach statistical significance in the participants analysis [$F_1(2,38) = 1.24$, $p > .05$, $\eta^2 = 0.061$], but it was significant in the items analysis [$F_2(2,18) = 4.27$, $p < .05$, $\eta^2 = 0.322$]. Nine dependent-samples t tests with a Bonferroni adjustment of alpha ($.05/9 = .0056$) were conducted to test pairwise differences among cell means within each level of article and sentence type. Within the article violation condition, pragmatically violated sentences yielded longer reading times than did control sentences. For sentences with the correct article, pragmatically violated sentences yielded longer reading times than did lexically violated or control sentences, which did not differ from each other. No other comparisons were significant.

The analysis of deviance ratings (see Table 3B) once again found a significant article \times sentence type interaction in both the participants and items analyses [$F_1(2,38) = 7.97$, $p < .005$, $\eta^2 = 0.296$; $F_2(2,18) = 3.87$, $p < .05$, $\eta^2 = 0.301$]. To test differences between cell means within each level, a series of nine dependent t tests with a Bonferroni alpha adjustment ($.05/9 = .0056$) was conducted. For sentences with article violations, lexically violated and pragmatically violated sentences did not differ from each other but both were rated more deviant than control sentences. The same pattern of results was found for sentences with correct articles. Within the control condition, sentences with article violations were rated as more deviant than sentences with correct articles. No other comparisons were significant.

A significant main effect of article was found in both analyses [$F_1(1,19) = 6.40$, $p < .05$, $\eta^2 = 0.252$; $F_2(1,9) = 29.87$, $p < .0005$, $\eta^2 = 0.768$], as sentences with article violations were given higher deviance ratings ($M = 4.64$, $SD = 1.93$) than were sentences with no violations ($M = 3.87$, $SD = 1.94$). Sentence type was also significant in both analyses [$F_1(2,38) = 55.51$, $p < .0001$, $\eta^2 = 0.745$; $F_2(2,18) = 45.61$, $p < .0001$, $\eta^2 = 0.835$]. Simple contrasts comparing the sentence type means collapsed over article (control, $M = 2.73$, $SD = 1.44$; lexically violated, $M = 4.74$, $SD = 1.58$; pragmatically violated, $M = 5.30$, $SD = 1.86$) showed all pairwise comparisons significant [control vs. lexically violated, $F(1,19) = 92.16$, $p < .0001$, $\eta^2 = 0.829$; control vs. pragmatically violated, $F(1,19) = 61.04$, $p < .0001$, $\eta^2 = 0.763$; lexically violated vs. pragmatically violated, $F(1,19) = 6.85$, $p < .05$, $\eta^2 = 0.265$].

Discussion

The overall reading times for Experiment 3 were much longer than those for either of the first two experiments. These reading times, recorded as they were within the context of the instruction to tell what might logically follow each sentence, suggest that in Experiment 3 we did in fact achieve the desired goal of ensuring that participants processed the sentences more deeply. Forcing meaningful semantic processing produced the same main effect of sentence type on reading times as that obtained in the German data in Experiment 1. Pragmatic violations thus took longer to read and identify than did lexical violations or control sentences, the same pattern as was obtained in Experiment 1. This suggests that the failure to have obtained this effect in Experiment 2 might have been due to participants' relatively more superficial reading of the sentences, which did not involve situation model processing to as great a degree as had occurred in Experiment 1.

On the other hand, Experiment 3 did not replicate the main effect of article violation on reading time found in

Experiment 1. The failure to obtain this effect was most likely due to the difference in linguistic depth of information carried in the German and English morphological violations. Whereas the German definite article carries information primarily about case, but also about gender, the English indefinite article carries only a change in sound harmony. Containing less information, this English indefinite article violation may have been a less diagnostic cue for the participants' purpose of making deviance ratings, precluding its use as an inferential shortcut for their task. This lack of diagnosticity may have resulted in the observed data; morphological violations did not reliably decrease reading times in Experiment 3.

GENERAL DISCUSSION

This series of experiments has demonstrated both some cross-linguistic generality and some need to recognize language-specific processing differences in sentence comprehension. In general, lexical and pragmatic violations slowed comprehension in both German and English. However, material must be processed to a fairly deep semantic degree for one to be able to observe this effect, a condition that may not have held in Experiment 2. The degree to which the morphological violations are relevant to deviance ratings seems to depend less on their surface syntactic and morphological constraints than on the depth of the linguistic information encoded by the morphological changes, at least under conditions in which participants can reasonably be expected to have been motivated to process the meaning of the sentences. We propose that the antecedents of such motivation may be found in part in recruitment procedure for the participants in Experiment 1 and in the instructions in Experiment 3.

The surface change of *den* to *der* in Experiment 1 and *a* to *an* in Experiments 2 and 3 involves a very similar sort of surface change in morphology, but the larger amount of information encoded in the German structure made it a more diagnostic cue for reaching participants' active goals of rating the sentences. As we have noted, this is probably due to the fact that the German definite article encodes several dimensions of grammatical information (case, gender, number), whereas the English indefinite article signals only number and whether the next word begins with a vowel or consonant. When one considers the effect of morphological changes, the amount of information carried and the complexity of the linguistic material encoded by these changes must be taken into account.

These results offer a cautionary note for psycholinguistic researchers. On the one hand, research involving materials from different languages is essential. Sometimes results replicate, and sometimes they do not, but one never knows without examining more than one language. The differential accessing of the propositional textbase and situation models levels in the lexical and pragmatic violation conditions appears to be fairly com-

parable in German and English, *when the comparability of the processing demands of the task is ensured.*

Second, and more subtly, manipulations that appear to be comparable across languages may not comparably affect processing. In the present experiments, the German definite article case violation in Experiment 1 and the English indefinite article sound harmony violation in Experiments 2 and 3 involved morphological changes that were structurally almost identical. However, the German article carries much more information and is integrally tied syntactically and semantically to more of the rest of the sentence than is the English violation. In Experiment 2, under conditions in which they were not explicitly required by task instructions to consider the sentence within its situation model context, participants may have been able to use the clear violation of language rules embodied in the English morphological violation to cut short situation model processing, leading to the lack of a difference between reading times for control and lexically violated and pragmatically violated sentences. As we stated previously, we believe that the control and lexical reading times did not differ because in both cases processing could stop with the propositional textbase. In Experiments 1 and 3, however, pragmatically violated sentences took longer to read because participants needed to reference the situation model level. In Experiment 2, it is probable that low participant motivation combined with the information-poor English morphological violation cues (*a/an*) to keep participants from engaging in more extensive processing of the stimulus sentences.

In sum, the various violation manipulations affected both the deviance rating and the more indirect reading time measures somewhat differently. Most clearly in Experiment 1, although pragmatically violated sentences were viewed the longest prior to rating, they were not rated as deviant as were lexically violated sentences, which were not viewed as long prior to rating but were rated as more deviant. Sentences that conflict only in the semantic interaction of two words (lexical violations) reference the propositional textbase level of representation and thus are easier to rate than those which require reference to the situation model level (pragmatic violations), which force consideration of the likelihood of various agents performing various actions in the real world.

During the initial parsing process, and during all of comprehension, information is being processed in parallel at multiple levels, including those of the surface code, propositional textbase, and situation model. However, the results of the present experiments suggest that the processing is not entirely parallel, with the situation model processing necessitating more time in a task requiring relatively deep semantic processing. There are many possible specific ways in which the posited parallel processing might occur. For example, Pinker (1999) presents a "words and rules" model whereby we store information both as words (e.g., individual lexical items, irregular verbs) and as rules (e.g., regular morphological rules).

The processing of *rules* (e.g., “add /s/, /z/, or /əz/ to make a noun plural”) taps into a different language module than does the processing of *words* (e.g., “*went* is the irregular past tense of *go*). In the present experiments the various violations involved both of these types. Thus the encountering of one of the morphologically violated lexical or pragmatic violations in the present research plugs into both of Pinker’s levels. How these two levels are consistent with Graesser et al.’s (1997) five levels of representation remains to be examined through future research.

REFERENCES

- BATES, E., & MACWHINNEY, B. (1987). Competition variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157-193). Hillsdale NJ: Erlbaum.
- CLAHSEN, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral & Brain Sciences*, **22**, 991-1013.
- CRAIN, S., & STEEDMAN, M. J. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 320-358). Cambridge: Cambridge University Press.
- CUETOS, F., & MITCHELL, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, **30**, 73-105.
- CUETOS, F., MITCHELL, D. C., & CORLEY, M. M. B. (1996). Parsing in different languages. In M. Carreiras, J. E. García-Albea, & N. Sebastián-Gallés (Eds.), *Language processing in Spanish* (pp. 145-187). Mahwah NJ: Erlbaum.
- FRAZIER, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 560-586). Hillsdale, NJ: Erlbaum.
- GILBOY, E., & SOPENA, J. M. (1996). Segmentation effects in the processing of complex NPs with relative clauses. In M. Carreiras, J. E. García-Albea, & N. Sebastián-Gallés (Eds.), *Language processing in Spanish* (pp. 191-206). Mahwah NJ: Erlbaum.
- GRAESSER, A. C., MILLIS, K. K., & ZWAAN, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, **48**, 163-189.
- INOUE, A., & FODOR, J. D. (1995). Information-paced parsing of Japanese. In R. Mazuka & N. Naga (Eds.), *Japanese sentence processing* (pp. 9-63). Hillsdale, NJ: Erlbaum.
- MAZUKA, R., & LUST, B. (1990). On parameter setting and parsing: Predictions for cross-linguistic differences in adult and child processing. In L. Frazier & J. de Valores (Eds.), *Language processing and language acquisition* (pp. 163-205). Dordrecht: Kluwer.
- MITCHELL, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601-617). Hillsdale, NJ: Erlbaum.
- PINKER, S. (1999). *Words and rules*. New York: Basic Books.
- PRITCHETT, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- ST. JOHN, M. F., & MCCLELLAND, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, **46**, 217-257.
- ZWAAN, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, **28**, 81-88.
- ZWAAN, R. A., & RADVANSKY, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, **123**, 162-185.

(Manuscript received September 15, 1998;
revision accepted for publication November 5, 2000.)