Journal of Psycholinguistic Research, Vol. 28, No. 6, 1999

Effect of Grammatical Gender and Semantic Context on Lexical Access in Italian

Simone Bentrovato,[1] Antonella Devescovi,[1] Simona D'Amico,[1] and Elizabeth Bates[2,3]

The interacting effects of sentence context and grammatical gender on lexical access were investigated in Italian using a timed picture-naming paradigm. Results showed large interacting effects of both sentence context and the gender of the article, with facilitation relative to two different control conditions. Repeat testing yielded an overall decrease in RT, but did not change the pattern of results. Results are interpreted in support of interactive activation models in which different sources of information are combined "on-line" to predict, anticipate or preactivate lexical targets.

Recent studies of Italian (Bates, Devescovi, Hernandez, & Pizzamiglio, 1996), French (Grosjean, Dommergues, Cornu, Guillelmon, & Besson 1994), German (Hillert & Bates, 1996; Jacobsen, this issue) and Russian (Akhutina, Kurgansky, Polinsky, & Bates, this issue) have shown that prenominal modifiers embedded in a short auditory phrase can prime the nouns they modify. Specifically, modifiers matching in grammatical gender can *facilitate* lexical access (decreasing reaction time relative to a neutral baseline) while modifiers

This research was completed in partial fulfillment of a university thesis (tesi di laurea) by the first author. It was supported by a grant to Elizabeth Bates, "Cross-linguistic studies in aphasia" (NIH/NIDCD RO1-DC00216). We are grateful for helpful comments and criticisms from Angela Friederici, Thomas Jacobsen, W. J. M. Levelt, Masha Polinsky, and other attendees at the July 1998 Workshop on Grammatical Gender in Leipzig, Germany (Max Planck Institute for Cognitive Neuropsychology), and for lengthy conversations about gender priming with Victor Ferreira, Kara Federmeier, Arturo Hernandez, Luigi Pizzamiglio, and Bernadette Schmitt. We also thank Nicole Wicha for her assistance in preparing these experiments, which are inspired by her research on gender and semantics in Spanish. None of these individuals are responsible for any errors or misinterpretations that might remain.

[1] University of Rome "La Sapienza," Rome.

[2] University of California, San Diego.

[3] Address all correspondence to Elizabeth Bates, Center for Research in Language 0526, University of California at San Diego, La Jolla, California 92093-0526.

0090-6905/99/1100-0677$16.00/0

with mismatching gender can *inhibit, suppress, or interfere* with lexical access (increasing reaction times relative to the same baseline). These results for gender are compatible with other forms of syntactic priming. For example, the time required to name or repeat a noun or verb can be facilitated or inhibited by simple syntactic contexts like "I like the ___ " or "I want to ___," relative to neutral contexts like "Now please say ___ " (Federmeier & Bates, 1997; Liu, 1996).

Although grammatical priming appears to be a robust phenomenon, its interpretation is controversial. Wright and Garrett (1984) were among the first to show that lexical decisions to target words are easier to make when those words are compatible with the grammatical context. Friederici and Schriefers have conducted several studies showing that reaction times to content words are slowed following a syntactic violation (Friederici & Schriefers, 1994; Schriefers, 1993). However, because their effects were largely inhibitory in nature, these authors concluded that grammatical priming is a postlexical phenomenon, affecting the time required to integrate target words into a phrase or sentence context (Jacobsen & Friederici, this issue). If this interpretation is correct, then the term "structural priming" may be a misnomer, because it does not involve a change in the process by which words are accessed in the first place.

To distinguish empirically between pre- and postlexical priming, Neely (1991) and Hernandez, Bates, & Avila. (1996) offer the following diagnostics: (1) prelexical priming effects are bottom-up and automatic, occurring in very short time windows, with minimal attention, impervious to strategies, characterized by facilitation rather than inhibition; (2) postlexical integration and/or prelexical guessing are both top-down and strategic, requiring long time windows between prime and target (e.g., SOA > 250 ms), influenced by instructions and experience in the task, and characterized primarily by inhibition. Using these diagnostics, Bates *et al.* concluded that gender priming does involve automatic prelexical activation: Their reaction times were very fast (<200 ms after the uniqueness point in a cued-shadowing task), with zero delay between prime and target. Because gender cues led to the right answer only 33% of the time, the best strategy in their experiment (if strategies were possible) would be to ignore the gender cue altogether, and yet participants seemed unable to suppress it. Finally, gender priming involved both facilitation and inhibition relative to gender-neutral adjectives (an ecologically valid baseline). Such facilitation is especially difficult to reconcile with postlexical integration accounts. By definition, postlexical integration can only occur after words have been accessed (O'Sheaghda, 1997). Furthermore, integration itself requires cognitive operations that cannot begin until lexical access is complete and presumably do not occur when words are accessed out of context or in a neutral context (e.g., ". . . .say _____ "). The logic is

straightforward: if operation A and operation B occur in a fixed serial order, then A + B cannot take less time than A alone. On these grounds, a post-lexical integration account cannot explain facilitative effects of grammatical context, assuming that we can find a neutral baseline that satisfies the above conditions.

In the present study, we will investigate the interaction between grammatical gender and sentence context in Italian. Using a design similar to a parallel study in Spanish by Wicha, Bates, Hernandez, Reyes, and Gavaldón de Barreto, (1998), we will show that gender and sentence meaning interact early in processing, affecting the time required to name pictures presented at unpredictable points within two-sentence auditory contexts. When gender and sentential semantics converge, reaction times are strongly facilitated relative to two different neutral baseline conditions. These results are compatible with interactive activation accounts in which different sources of information are combined, quickly and in parallel, to predict the identity of an upcoming word (Allopenna, Magnuson, & Tanenhaus, 1998; Elman, 1990; MacDonald, Pearlmutter, & Seidenberg, 1994; MacWhinney & Bates, 1989; Marslen-Wilson & Tyler, 1987; Van Petten, Coulson, Rubin, Plante, & Parks, 1999).

METHOD

Participants

A total of 25 native speakers of Italian (13 women) participated in the main experiment; 13 (7 women) returned approximately one week later for a repeat testing session (all participants were asked to return; only 13 were able to do so). Fifty additional native speakers (25 women) participated in a separate study to obtain picture-naming times for a larger list of items, presented out of context. All participants were undergraduate students of the University of Rome "La Sapienza" Department of Psychology, and were paid US $5 for their participation. The average age for participants in the main experiment was 23.1 years (range = 20–27). The average age for the norming study was 24.3 years (range = 19–30).

Materials

Every stimulus comprised a brief two-sentence discourse in the Italian language, presented auditorily, and a target picture presented at some point within one of the two sentences (target positions were varied within and between sentences).

Picture Targets

A total of 110 picture targets were used, each viewed and named only once by any individual participant. All were line drawings of common objects (Abbate & LaChappelle, 1984; Dunn & Dunn, 1981; Goodglass, Kaplan, & Weintraub, 1983; Snodgrass & Vanderwart, 1989), chosen from a larger list of 520 pictures from an international picture-norming study (Bates *et al.*, research in progress, October 1998). Based on results for 50 native speakers of Italian (see above), pictures for the main experiment were selected on the following criteria: A single name was provided by at least 80% of the norming participants, no more than four different names were provided by these participants, the primary name for the target is not plural, does not begin with a vowel or a fricative consonant, and does not belong to a small class of exception words (e.g., words like *mano* (for *hand*), which is feminine in Italian but ends in "-o", the typical ending for masculine words).

The 110 picture names selected on these criteria break down as follows: 56 masculine (44 ending with the phonologically transparent masculine vowel "-o", 12 with the phonologically ambiguous vowel "-e") and 54 feminine (46 ending with the phonologically transparent feminine vowel "-a", 8 with the phonologically ambiguous vowel "-e"). To obtain an approximate but representative balance by noun gender and suffix type, we had to ease our criteria for two words: *Cassaforte* (which means safe or strongbox) elicited a total of six different names, and *botte* (which means keg), which was named by only 76.6% of participants in the naming study. The mean length in letters of the 110 picture names was 6.5, with a range of 4–11 (note that words tend to be substantially longer in Italian than English); mean length in syllables averaged 2.7, with a range of 2–4. Most pictures represented inanimate objects that are frequently encountered in everyday life, including household objects, articles of clothing, or musical instruments; 6 figures represented parts of the human body or human characters [e.g., *regina* (queen)], while another 15 were animals with a fixed grammatical gender that was independent of the animal's biological sex (e.g., *tigre*—a feminine noun meaning tiger; *ragno*—a masculine noun meaning spider). For a few of these animal names, both a masculine and a feminine noun is available in the language, depending on the animal's biological sex, but one form is far more common and represents the citation form [e.g., cavallo (horse) can be replaced by the alternative *cavalla* when referring explicitly to a female of the species, equivalent in use and frequency to the English counterpart *mare*]. There were, in fact, no cases in the picture-norming experiment in which the low-frequency alternative was provided for these pictures.

Within each sentence stimulus, the target picture appeared immediately after the article (zero interval from offset of the auditory article to onset of the picture).

Auditory Sentence Contexts

A total of 550 two-sentence contexts were prepared, 440 for use within the factorial gender-by-semantics design, and another 110 neutral controls designed to work equally well with any target picture, containing no semantic or gender constraints. The 440 sentence contexts represent four variations each of 110 contexts designed to favor production of one target picture name, with the target placed in either the first or the second sentence. For example, for the target picture/name *book,* the following context was generated (where the target picture with its target name is represented in capital letters):

> Quando vado a letto prima di addormentarmi leggo sempre un *libro.* Per questo mia mamma mi ha regalato una collezione di romanzi gialli.
>
> When I go to bed before falling asleep I always read a *book.* For this reason, my mother gave me a collection of murder mysteries.

In order to create a gender-incongruent version of the same stimulus, the article before the target was switched to the opposite gender, e.g., from *un* to *una* in the above example, as follows:

> Quando vado a letto prima di addormentarmi leggo sempre una *libro*. . . .

For approximately half the items, the article before the target was indefinite (*un* in masculine, *una* in feminine, as in the above example). For the remaining items, the article before the noun was definite (*il* in masculine, *la* in feminine).

Once the 110 base sentences were created (with their 110 gender-incongruent variants), semantically incongruent versions were assembled by selecting from the 110 target pictures an item of the same gender that was a highly implausible completion on semantic grounds (these assignments were made on subjective grounds, rather than through random assignment, in order to ensure that a similar degree of implausibility would result across items). For example, the above sentence context designed to elicit *libro* (or book) was coupled with the picture *topo* (mouse) in the target position, as in

> Quando vado a letto prima di addormentarmi leggo sempre un *topo*. . . .
>
> When I go to bed before falling asleep I always read a *mouse*. . . .

The fourth and final version of each sentence context was created by switching the gender of the article before the implausible alternative, as in

> Quando vado a letto prima di addormentarmi leggo sempre una *topo*. . . .

We will refer to the four versions of each stimulus with the following symbols: +G+S refers to items that are congruent in both gender and seman-

tics, –G+S to items that are incongruent in gender only, +G–S to items that are incongruent in semantics only, and –G–S to items that are incongruent along both dimensions.

In addition to the 440 experimental items (four each for 110 pictures), we also developed a set of 110 sentence contexts (following Wicha *et al.*) that were equivalent in length and complexity to the experimental contexts without providing relevant gender or semantic constraints. An example would be the following:

> Silvia ha fatto un test in inglese in cui doveva ripetere *libro* per cinque volte. L'insegnante ha detto che è l'unico modo per migliorare la pronuncia.
>
> Sylvia took a test in English in which she had to repeat *book* five times. The teacher said that was the only way to improve her pronunciation.

These 110 contexts served as the within-experiment neutral baseline, used to assess whether any of the gender-by-semantics conditions resulted in facilitation or inhibition of the target picture name. (A complete list of items can be provided by the authors on request.)

List Construction

In order to ensure that no single participant ever heard the same sentence or saw the same picture twice in the course of the experiment, five lists were prepared. In the first list, the 110 pictures were assigned quasi-randomly to one of the five experimental conditions (+G+S, –G+S, +G–S, –G–S, neutral control), until five conditions with 22 items each were obtained, with no repetitions of pictures or their associated sentences. Items were then rotated across the five conditions, one condition per list (e.g., *book* might occur in –G+S in list 1, +G+S in list 2, +G–S in list 3, –G–S in list 4, and neutral in list 5).

Procedure

Auditory stimuli for the main experiment were digitally recorded in a soundproof chamber and transferred onto a Macintosh computer as individual SoundEdit 1.0 files. Sentences were cleaned (blank spaces before and after the utterance removed) and sound labels were placed before and after the location of the target. A time-lock marker was placed at the end of the first part (i.e., before the target). Once this process was completed, they were converted into individual PsyScope files (Cohen, MacWhinney, Flatt, & Provost, 1993) for experimental presentation. The average length of the auditory contexts (excluding the interval in which pictures were presented) was 11.24 s (range = 8.13–15.37 s).

The picture targets were digitized images set in black outline on a white background. Individual files were created for each image, so that they could be called into the appropriate sentence during on-line presentation. The time between offset of the auditory context and onset of the picture was placed at zero.

Testing

Participants were tested one at a time in a quiet cubicle. Stimuli were presented on a Macintosh Performa 6214CD, using PsyScope presentation software (Cohen *et al.,* 1993). Participants wore headphones with adjustable volume that were connected to the sound amplifier port of the Performa. The headset had a sensitive, built-in microphone that was connected to the Carnegie Mellon Button Box, a measuring device with 1-ms resolution designed for use with Macintosh computers. Response times were collected in milliseconds using the CMU button box, which was connected to the Performa modem port. The experimenter also wore headphones (connected to the Performa via a two-prong connector), and hand-recorded all naming errors on a score sheet during testing.

Participants were randomly assigned to one of the five lists based on their participant number. At the beginning of each session, a summarized version of the instructions appeared on the screen. The experiment began when the participant pressed a button on the keyboard. The sentences were randomly presented in a continuous sequence. During the auditory presentation, a fixation point (++) appeared in the center of the screen. At some point in the sentence, a picture appeared on the screen in place of the fixation point, and the auditory presentation was halted. Participants were instructed to name the object as quickly as possible upon presentation. The image remained on the screen for 5 s, or until the participant responded, whichever came first. An "NR" for no response was marked in the data file if the image disappeared prior to the participant's response. The image disappeared as soon as the microphone picked up the onset of a response, followed by a 500-ms delay, then the auditory stimulus continued accompanied by a blank screen. There was a 3-s delay between the end of each sentence and initiation of the next trial.

Prior to the main experiment, participants were given a brief practice session comprising ten stimuli (structurally similar but different in content from those used in the main experiment), two each from each of the five conditions that would be encountered in the main experiment. They were instructed to speak directly into the microphone, as clearly as possible, to name the picture with a single word (if possible), and not to emit any other sounds (no clearing of the throat, no preparatory sounds like "uhmmm," etc.). There was a

pause after half the sentence pairs had been presented. Participants were allowed to continue when ready, by pressing any button on the keyboard. The entire experimental session lasted approximately 45 min.

These procedures differ in several respects from the parallel study of Spanish by Wicha *et al.* First, new materials were developed for Italian (i.e., stimuli are not translations). Second, we used a 5000-ms response window (vs. 3000 ms. in Wicha *et al.*). Third, to ensure that participants were attending to the sentence contexts, Wicha *et al.* administered occasional multiple-choice questions, presented visually at random intervals on approximately 10% of trials, asking about the content of the immediately preceding discourse. We omitted this procedure to learn more about the effects of context when listeners have the option of ignoring sentence context (a strategy that would permit them to avoid errors on the 60% of items that are discordant in gender, semantics, or both). Finally, we retested half the participants (13 of 25), one week after the main experiment, to determine whether strategies would emerge with repeat testing that modify the shape of priming effects. The second session (time 2) was identical to the first, with participants assigned to the same stimulus list they had received at time 1. However, the order of presentation of stimuli within the list was randomized (as it was in the first session), so that the actual stimulus order varied between time 1 and time 2.

Picture Norming

A separate set of 50 subjects had participated several months earlier in a norming study designed to assess name agreement and naming times for 520 pictures of common objects, including the 110 that were chosen for the main experiment. In this study, pictures were presented to each participant in one of ten randomized orders (five participants per list). The experimental apparatus, mode of presentation, recording, and checking of errors were identical to those adopted in the present experiment (above). At the beginning of each trial, a fixation cross appeared in the middle of the screen; 500 ms after the appearance of the cross, a brief pure tone sounded. Immediately after the offset of the tone, the picture appeared. The interval between trials was 3000 ms. If no response was given within 4000 ms from picture onset, the trial ended and the next trial began. If the participant responded, the picture disappeared as soon as the response was detected by the CMU button box; an additional 500 ms passed in order to record the response without interruption, and then the trial was ended and the intertrial interval began. Although a number of dependent variables were used in the norming study, we based our comparisons here on mean reaction times for those trials in which participants in the norming study produced the target word (subject to the same data-cleaning criteria described below).

RESULTS AND DISCUSSION

Errors

Of the trials, 2.58% were eliminated because the participant failed to respond, made an irrelevant noise before responding, or the voice key failed to register the response. An additional 3.3% were eliminated because participants produced an alternative response, including frank semantic errors (1.6%) or semantically correct synonyms (1.7%) that changed the gender of the noun and thereby turned a nonmatching gender condition into a matching one. For example, a speaker might produce *il barile* (masculine *keg*) when *la botte* (feminine *barrel*) was the target word produced by at least 80% of respondents in the picture-norming task. Reaction time analyses are based on the remaining 94.22% of trials with a correct response.

Because error rates were so low, they were not subjected to analyses of variance. However, the small category of semantically correct substitutions with a gender switch are of theoretical interest [e.g., barrel (feminine) for keg (masculine)], because these represent a solution to the problem posed by gender discordance. Across the five experimental conditions, error rates for such substitutions broke down as follows: neutral, 0.7%; +G+S, 0.36%; +G–S, 0.72%; –G+S, 3.27%; –G–S, 3.45%. A chi-square analysis over these errors (collapsed over subjects) showed that the distribution across conditions deviated from chance ($p < .001$). Hence, although such substitutions are rare, they indicate that some participants were using a combination of gender context and pictorial information to outwit the experimenters on at least a few items.

Response Times

Response times (RT) were first subjected to a 2 (matching gender, nonmatching gender) by 2 (semantically congruent, semantically incongruent) within-subjects analysis of variance, over participants (F_1) and items (F_2). Comparisons with the neutral-control condition were reserved for *post hoc* analyses. Collapsed across the four experimental conditions, the mean RT was 819 ms (SD = 149.4 ms, SE = 14.93 ms). Despite differences in language, procedure, and materials, results are close to the 855 ms mean reported by Wicha *et al.*

In the 2 × 2 analysis, there was a significant main effect of gender, reflecting faster RTs in the matching-gender conditions [F_1 (1, 24) = 32.03, $p < .0001$; F_2 (1, 109) = 55.99, $p < .0001$]. There was also a significant main effect of semantics, with faster RTs in the semantically congruent conditions [F_1 (1, 24) = 13.30, $p < .001$; F_2 (1, 109) = 43.89, $p < 0.0001$]. The

interaction between gender and semantics was significant [F_1 (1, 24) = 13.59, $p < .001$; F_2 (1, 109) = 20.03, $p < .0001$]. Table I presents the mean RTs, *SD*s and *SE*s for each condition Figure 1 displays the interaction between gender and semantics (where the dotted lines indicate performance in the within-experiment neutral-control condition and the out-of-context picture-norming control condition, both discussed below).

To explore the gender × semantics interaction further, simple one-way F tests were run comparing each of the conditions to the others, over subjects and items. In each relevant comparison, the fully congruent condition (+G+S) was significantly faster than the other three [with +G–S, F_1 (1, 24) = 22.30, $p < .0001$, F_2 (1, 109) = 39.87, $p < .0001$; with –G+S, F_1 (1, 24) = 76.91, $p < .0001$, F_2 (1, 109) = 110.8, $p < .0001$; with –G–S, F_1 (1, 24) = 53.16, $p < .0001$, F_2 (1, 109) = 73.62, $p < .0001$]. However, the three conditions that contained some form of incongruence did not differ significantly from one another [+G–S with –G+S, F_1 (1, 24) = 0.11 n.s., F_2 (1, 109) = 0.27 n.s.; +G–S with –G–S, F_1 (1, 24) = 1.97, n.s. F_2 (1, 109) = 3.47 n.s.; –G+S with –G–S, F_1 (1, 24) = 1.16 n.s., F_2 (1, 109) = 1.44 n.s.].

The broad finding is that congruent materials are faster than stimuli containing one or more discrepancies (gender, semantics, or both). To determine whether these results reflect facilitation or inhibition, each of the four conditions was compared with performance in the neutral sentence contexts, again using simple one-way analyses of variance over both participants and items. Responses significantly faster than baseline reflect facilitation, and responses significantly slower than baseline reflect inhibition.

There was robust evidence for contextual facilitation in the fully congruent condition (+G+S), compared to the neutral baseline [F_1 (1, 24) = 53.03, $p < .0001$; F_2 (1, 109) = 79.28, $p < .0001$]. As Table 1 shows, this facilitative effect is very large, reflecting a mean RT difference of 126 ms. There was also evidence for contextual inhibition in the condition where the target was maximally incongruent (–G–S), compared to the same neutral baseline [F_1 (1, 24) = 4.88, $p < 0.037$; F_2 (1, 109) = 4, $p < 0.048$]. However, this inhibitory effect is considerably smaller than the facilitative effect, reflecting a mean RT

Table I. Performance in Main Experiment (Gender × Semantics)

Condition	+G+S	+G–S	–G+S	–G–S	Neutral
Mean RT	703	841	852	879	829
Standard deviation	106	146	146	136	131
Standard error	21	29	29	27	26
Total semantic errors	4	9	8	10	13
Total semantic synonyms	2	4	18	19	4
Total alternative responses	6	13	26	29	17

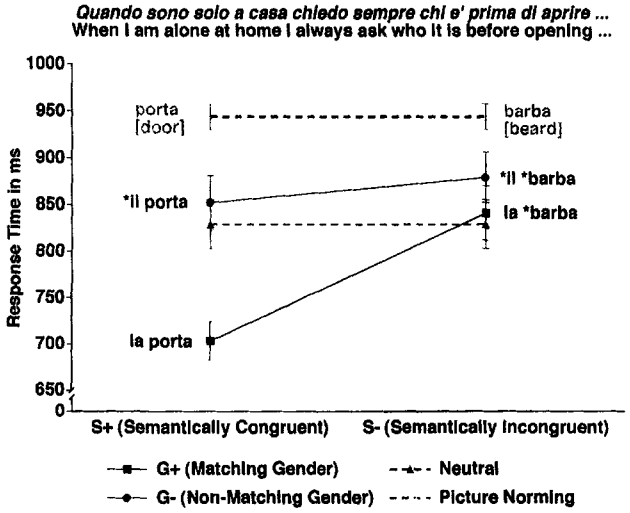

Fig. 1. Interaction for picture-naming task between gender and semantic priming in milliseconds, with standard error bars. Dotted lines reflect two neutral baselines: neutral sentences (within-subjects, lower line) and naming out of context (between-subjects, upper line).

difference of 50 ms. The two remaining conditions (in which one source of information is congruent while the other is not) did not differ significantly from the neutral baseline, thus providing no evidence for inhibition or facilitation [+G–S vs. baseline, F_1 (1, 24) = 0.27, n.s. F_2 (1, 109) = 0.16 n.s.; –G+S vs. baseline, F_1 (1, 24) = 1.20, n.s. F_2 (1, 109) = 2.05 n.s.].

To summarize, gender and semantics do interact to affect response times in a picture-naming paradigm. This interaction involves both a large degree of facilitation (when both sources of information converge) and a smaller effect of inhibition (when both sources of information lead to the wrong answer). When only one dimension is discordant with the picture target, responses do not differ from the neutral baseline (which provides no gender or semantic constraints), suggesting that speakers in this condition have to "start from scratch" when their expectations are disconfirmed. These results constitute a full replication of Wicha *et al.*, despite differences in language, materials, and certain procedural details (i.e., the absence in our study of multiple-choice questions to assure attention to the context).

Comparisons with Picture-Norming Study

We designed the neutral stimuli to ensure that sentence contexts were matched for length and complexity, so that the baseline condition did not

differ from experimental stimuli in processing load, orientation of attention, and other cues that might affect response preparation. However, there is no such thing as a perfect baseline, particularly once we move from a set of crosses in the middle of the screen to something that is supposed to serve as a "neutral paragraph." For example, it could be argued that our neutral condition is harder than the four experimental conditions because the latter provide a greater degree of syntactic constraint (i.e., because articles are usually followed by nouns in Italian, participants can begin to prepare their naming response as soon as an article is encountered, whether or not the article contains the right gender). One might also argue that these baseline contexts are difficult to process on other grounds, involving metalinguistic situations that are fairly rare in natural discourse (". . . .repeat *book*"). If our baseline stimuli are somehow more difficult to process than the experimental materials (despite the many incongruencies that the latter contain), then we may have overestimated the magnitude of contextual facilitation and underestimated response inhibition in the present study.

In response to this concern, the five within-subject conditions in the present experiment were compared with responses by subjects who encountered the same 110 picture stimuli out of context, embedded among the 520 stimuli employed in a picture-norming study. The norming study provides information about the time required to name these pictures when there is no grammatical or semantic information of any kind, other than the within-experiment expectation that all items will be names for common objects and hence nouns. However, the norming experiment does provide nonverbal visual and auditory cues to picture onset, reducing response preparation time.

The mean RTs obtained in the picture-norming study for the 110 picture targets used here was 944 ms (SD = 144.6, SE = 13.8). This mean RT is plotted in Fig. 1, represented as a flat dotted line crossing performance in the four experimental conditions (parallel to the dotted line representing the mean RT of 829 ms in the experimental-internal neutral-control condition). Although one must be cautious in interpreting absolute differences in reaction time when two different groups of participants are compared, it should be clear from Fig. 1 that naming out of context is more difficult (i.e., it takes more time) than naming of the same pictures in any of the five conditions employed in the present experiment. Note also that participants in the norming study were drawn from the same pool (undergraduate students at the University of Rome), in the same age range, paid the same amount for their participation, and they were tested in the same experimental apparatus under similar conditions of picture presentation. Thus there is every reason to expect comparable response times. Yet the difference between our experiment-internal neutral baseline and performance out of context was 115 ms, a substantial savings. Indeed, performance in the norming study

was slower than performance in the worst condition in the main experiment (–G–S), by approximately 66 ms. We suggest that performance was faster in the main experiment because some linguistic context is better than none at all, even if that context contains misleading cues, due to some (still unknown) combination of prosodic and grammatical information that facilitates noun retrieval in the picture-naming task.

Retest

The purpose of the retest manipulation was to determine whether, with practice and repetition, participants could develop strategies to optimize performance on this task. Recall that 60% of items contain an anomaly of some kind, and only 20% have contexts that are fully congruent with and predictive of the target name. Under such conditions, participants might decide to ignore the context altogether in order to minimize error and maximize efficiency. This strategy was precluded by Wicha *et al.*, who required participants to answer occasional multiple-choice questions, but it would be possible in the present study. A general strategy of ignoring the auditory context should (if it were possible to do so) tend to reduce priming effects across the board, including the large facilitative effects reported above.

Reaction times for session 1 and session 2 were calculated for the 13 participants who attended both sessions, using the same criteria for error coding and data cleaning described above. The resulting RTs (for correct responses only) were subjected to two analyses of variance. The first analysis excluded the neutral baseline condition, comprising a 2 (test vs. retest) × 2 (congruent vs. incongruent gender) × 2 (congruent vs. incongruent semantics) design. The second analysis included the neutral baseline condition, within a 2 × 5 within-subject design, with two levels for session (test vs. retest) × five levels of stimulus condition (+G+S, +G–S, –G+S, –G–S, neutral) design. The point of this second analysis was to determine whether the global pattern of results remained the same across sessions, without repeating a large number of two-way *post hoc* comparisons with the neutral baseline.

The first analysis yielded a large main effect of session [F_1 (1, 24) = 13.96, $p < .0001$; F_2 (1, 109) = 155.9, $p < .0001$], reflecting a drop in mean RT of approximately 124 ms from session 1 (mean RT = 852 ms, $SD = 143$, $SE = 20$) to session 2 (mean RT = 721 ms, $SD = 91$, $SE = 13$). There were also significant main effects of gender [F_1 (1, 24) = 57.34, $p < .0001$; F_2 (1, 109) = 58.86, $p < .0001$] and semantics [F_1 (1, 24) = 22.74, $p < .0001$, F_2 (1, 109) = 46.280, $p < .0001$], plus a significant gender by semantics interaction [F_1 (1, 24) = 15.50, $p < 0.001$; F_2 (1, 109) = 21.187, $p < .0001$]. The gender and semantics effects were all in the same direction reported for the main experiment. Most important for our purposes here, there were no sig-

nificant interactions of session with either of the contextual variables (all $F < 2.0$, $p > .15$; all $F_2 < 1.5$, $p > .13$). Hence, even though participants were able to profit markedly from stimulus repetition in their overall reaction times, the pattern of response did not change.

The second analysis pushed the issue a step further by including the neutral-control conditions at sessions 1 and 2, combined into a single five-level variable. There was (again) a significant main effect of session [F_1 (1, 24) = 12.91, $p < .001$; F_2 (1, 109) = 181.2, $p < .0001$], reflecting an overall drop of 124 ms collapsed across contextual conditions (session 1, mean RT = 853, SD = 141, SE = 12; session 2, mean RT = 729, SD = 85, SE = 11). There was also a significant effect of context [F_1 (4, 96) = 24.96, $p < .0001$, F_2 (4, 436) = 33.42, $p < .001$]. However, the interaction between session and context did not reach significance [F_1 (4, 96) = 1.36, $p > .25$; F_2 (4, 436) = 1.923 $p > .10$]. In other words, the shape of the context effect (including facilitation) does not change over sessions, even though RTs are substantially faster on retest. It seems that participants in this experiment do not develop a strategy to ignore the sentence context, even with repeat testing. Such imperviousness to strategies is what we would expect if priming effects are automatic, robust, and difficult to suppress (Den Heyer, Goring, & Dannebring, 1985; Durgunoglu, 1988).

CONCLUSION

The purpose of this study was to investigate the interacting effects of grammatical gender and sentential context on word retrieval and production in a timed picture-naming paradigm. It parallels a study in Spanish by Wicha *et al.* (1997), and in every key respect the two studies yield comparable results, despite differences in language, stimulus materials, and small but potentially important variations in method (e.g., presence/absence of multiple-choice questions to force attention to the sentence context). Both studies revealed large and significant interactions between sentence context and grammatical gender, due primarily to response facilitation when both sources of information converge. When only one dimension is incongruent (semantics or gender), reaction times are no different from baseline, as if these speaker/listeners had no prior information of any kind (forced to flush the system and start from scratch). When both dimensions are violated, a small but significant degree of response inhibition is observed, relative to the same baseline.

Following the diagnostics proposed by Neely (1991), Hernandez *et al.* (1996) and Bates *et al.* (1996), we conclude that the interacting effects of gender and sentence meaning in this experiment reflect automatic, prelexical lexical priming. RTs were relatively fast (with Italian words that average

2.7 syllables in length), with a minimal SOA (zero interval between context offset and target onset), in a picture-naming task that requires no postlexical decisions of any kind. Although the most efficient strategy in this experiment would be to ignore contexts that pay off on only 20% of trials and are frankly misleading 60% of the time, it seems that participants are unable to suppress this information, even with repeat testing, and without multiple-choice questions that force attention to the auditory sentence contexts. Finally, the convergence of gender and sentence meaning led to massive response facilitation relative to two different neutral baselines. If gender priming reflects postlexical integration, as some investigators have argued, it should not be possible to obtain faster RTs with both lexical access and integration (A + B) than we can observe when lexical access is measured out of context, or in a neutral context that requires no integration (A only). Hence response facilitation provides strong evidence against the postlexical account.

In defense of the postlexical account, one might argue that neither of our baselines were appropriate. Like Wicha *et al.*, we developed linguistically neutral stimulus materials that offered no gender or semantic constraints, but matched the experimental stimuli in terms of length, complexity, processing load, attention orientation, and response preparation. We are convinced that this is an ecologically valid baseline, as close as one can get to the challenging notion of a "neutral paragraph," relatively free of both semantic and grammatical constraints but preserving the sound and flow of normal auditory input. Certainly it is preferable to (for example) scrambled strings of unrelated words, or content-free Jabberwocky (for a discussion of such control conditions, see Marslen-Wilson & Tyler, 1981; Van Petten *et al.*, 1999). However, some criticisms have been raised regarding the frequency and "naturalness" of these metalinguistic contexts (e.g., "Armando had to repeat *book* five times. . . ."). One might argue that all of our materials except for the fully congruent condition are difficult and "odd," providing a misleading view of the proportional contributions of facilitation and inhibition within this experiment. To deal with this criticism, we compared responses in all five experimental conditions with RTs for an independent sample of speakers who named the same pictures out of context, embedded within a larger list of 520 items. Although this is a between-subjects comparison (so that absolute reaction times should be interpreted with caution), participants in the two studies were drawn from the same population and tested under highly similar conditions. If anything, one would assume that college students asked to name 520 pictures in a row would be anxious to escape from the situation, responding as quickly as possible and picking up speed across the course of the experimental session. Results were very clear: Reaction times in the norming study for the same 110 pictures were slower than reaction times in any of the linguistic conditions used in

our experiment. We suggest that any linguistic context that is at all natural (auditory presentation, normal prosody, familiar lexical items in familiar grammatical structures) is better than no context at all, even though subjects in both experiments were aided by nonlinguistic orienting cues (fixation cross in both experiments; tones in the picture-norming study). Based on this finding, we conclude that the neutral baseline used in our experiment is valid, and the facilitative effects observed against this baseline represent a real savings in response time due to the presence of converging grammatical and semantic information.

These results add to a growing body of evidence for anticipatory, predictive effects of context on word and sentence processing, in both comprehension and production (e.g., Allopenna *et al.,* 1998; Altmann, van Nice, Garnham, & Henstra, 1998; MacWhinney & Bates, 1989; Marslen-Wilson & Tyler, 1981, 1987; MacDonald *et al.,* 1994; Van Petten et al., 1999). Although it has been argued that such anticipatory effects would be exceedingly expensive on computational grounds (O'Sheaghda, 1997), such costs are not incurred in a neural network in which lexical access involves parallel computation in a high-dimensional space of sound and meaning (e.g., Elman, 1990). Hence, on both logical and empirical grounds, it is possible to entertain a model in which grammatical and lexical information are used together, early in processing, to predict and anticipate upcoming words and structures.

REFERENCES

Abbate, M. S., & LaChapelle, N. B. (1984). *Pictures, please!* Communication Skill Builders, Inc.

Akhutina, T., Kurgansky, A., Polinsky, M., & Bates, E. (this issue). Processing of grammatical gender in a three-gender system: Experimental evidence from Russian. *Journal of Psycholinguistic Research.*

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J.-A. (1998). Late closure in context. *Journal of Memory and Language, 38,* 459–484.

Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics, 85*(7), 992–1004.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments & Computers, 25,* 257–271.

Den Heyer, K., Goring, A., & Dannebring, G. L. (1985). Semantic priming and word repetition: The two effects are additive: *Journal of Memory and Language, 24,* 699–716.

Dunn, Lloyd M., & Dunn, Leota M. (1981). *Peabody Picture Vocabulary Test - Revised.* Circle Pines, MN: American Guidance Service.

Durgunoglu, A. Y. (1988). Repetition, semantic priming, and stimulus quality: Implications for the interactive-compensatory reading model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 590–603.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211.

Federmeier, K. D., & Bates, E. (1997). Contexts that pack a punch: Lexical class priming of picture naming. *Center for Research in Language Newsletter, 11*(2). La Jolla: University of California, San Diego.

Friederici, A. D., & Schriefers, H. (1994). The nature of semantic and morphosyntactic context effects on word recognition in young healthy and aphasic adults. In D. Hillert (Ed.). *Linguistische Berichte: Linguistics and Cognitive Neuroscience: Theoretical and empirical studies on language disorders.* Germany: Westdeutscher Verlag.

Goodglass, E., Kaplan, E., & Weintraub S. (1983). *The Boston Naming Test.* Philadelphia: Lee & Febiger.

Grosjean, F., Dommergues, J-Y, Cornu, E., Guillelmon, D., & Besson, C. (1994). The gender-marking effect in spoken word recognition. *Perception & Psychophysics, 56*(5), 590–598.

Hernandez, A., Bates, E., & Avila, L. X. (1996). Processing across the language boundary: A cross-modal priming study of Spanish-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(4), 846–864.

Hillert, D., & Bates, E. (1996). *Morphological constraints on lexical access: Gender priming in German* (Tech. rep. No. 9601). La Jolla: University of California, San Diego, Center for Research in Language

Jacobsen, T. (1999). Effects of grammatical gender on picture and word naming: Evidence from German. *Journal of Psycholinguistic Research, 28(5),* 499–514.

Jacobsen, T., & Friederici, A. (this issue). Introduction.

Liu, H. (1996). *Lexical access and processing dissociations in nouns and verbs in a second language.* Doctoral dissertation, University of California, San Diego.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703.

MacWhinney, B., & Bates, E. (Eds.) (1989). *The crosslinguistic study of sentence processing.* New York: Cambridge University Press.

Marslen-Wilson, W. D., & Tyler, L. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society, B,* 317–332.

Marslen-Wilson, W. D., & Tyler, L. K. (1987). Against modularity. In J. Garfield (Ed.), *Modularity and knowledge representation.* Cambridge, MA: MIT Press

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

O'Seaghdha, P. (1997). Conjoint and dissociable effects of syntactic and semantic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 807–828.

Schriefers, H. (1993). Syntactic process in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 841–850.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Timecourse of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory and Cognition* (in press).

Wicha, N. Y., Bates, A., Hernandez, A., Reyes, I., & Gavaldón de Barreto, L. G. (1997). *When zebras become painted donkeys: The interplay between gender and semantic priming in a Spanish sentence context* (Technical Report 9701). La Jolla: University of California, San Diego, Center for Research in Language.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences. *Memory & Cognition, 12,* 31–45.