

The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect

Lin Wang^{a,b,c}, Marcel Bastiaansen^{b,c}, Yufang Yang^{a,*}, Peter Hagoort^{b,c,**}

^a Institute of Psychology, Chinese Academy of Sciences, Datun Road 4A, Beijing 100101, China

^b Max Planck Institute for Psycholinguistics, The Netherlands

^c Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands

ARTICLE INFO

Article history:

Received 17 March 2010
Received in revised form
14 September 2010
Accepted 22 December 2010
Available online 30 December 2010

Keywords:

Information structure (IS)
Wh-question–answer pairs
Pitch accent
Semantic illusion
N400 effect
Semantic integration

ABSTRACT

To highlight relevant information in dialogues, both wh-question context and pitch accent in answers can be used, such that focused information gains more attention and is processed more elaborately. To evaluate the relative influence of context and pitch accent on the depth of semantic processing, we measured event-related potentials (ERPs) to auditorily presented wh-question–answer pairs. A semantically incongruent word in the answer occurred either in focus or in non-focus position as determined by the context, and this word was either accented or unaccented.

Semantic incongruency elicited different N400 effects in different conditions. The largest N400 effect was found when the question-marked focus was accented, while the other three conditions elicited smaller N400 effects. The results suggest that context and accentuation interact. Thus accented focused words were processed more deeply compared to conditions where focus and accentuation mismatched, or when the new information had no marking. In addition, there seems to be sex differences in the depth of semantic processing. For males, a significant N400 effect was observed only when the question-marked focus was accented, reduced N400 effects were found in the other dialogues. In contrast, females produced similar N400 effects in all the conditions. These results suggest that regardless of external cues, females tend to engage in more elaborate semantic processing compared to males.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Conversation is one of the most common ways of communication in social life. Whether consciously or unconsciously, people generally construct their utterances by following particular rules. Take the following dialogue as an example: Speaker A: *What kind of vegetable did mum buy to cook?* Speaker B: *Mum bought EGGPLANT to cook.* Speaker A brings up a topic and opens a slot for speaker B to fill in, and speaker B provides the relevant information (the word in capitals) and emphasizes it through accentuation to make communication more efficient. Therefore, the information put forward by speaker A becomes common ground between the interlocutors, whereas the information provided by speaker B (*eggplant*) is highlighted relative to the shared knowledge. This way of linking novel information with given information is termed information

structure (IS; Jackendoff, 2002). Generally speaking, IS contains two elements: background and focus. The background is usually shared information between the interlocutors, while the focus is the new or contrastive information that is salient to them (Günther, Maienborn, & Schopp, 1999). Two prominent means by which such salient information is placed in focus are (i) the context of the ongoing discourse (such as a wh-question in a question–answer pair), and (ii) prosodic features (for Dutch mainly pitch accent). For the sake of clarity, we will refer to requested and background information as “focus” and “non-focus”, respectively, and to the information marked by prosodic features as “accented” and “unaccented” information.

Behavioral studies have shown that in languages such as English and Dutch language comprehension is faster and easier when focused information is accented and non-focused information is unaccented (Bock & Mazzella, 1983; Dahan, Tanenhaus, & Chambers, 2002; Terken & Nootboom, 1987). Moreover, focused information receives more attention and is processed more deeply (Birch & Rayner, 1997; Cutler & Fodor, 1979; Ward & Sturt, 2007).

Several ERP studies have investigated temporal aspects of the influence of IS on language comprehension, in different languages. However, the results are not straightforward. For missing accentuation (i.e. focus information that is unaccented), most studies

* Corresponding author.

Tel.: +86 10 64888629; fax: +86 10 64872070.

** Corresponding author at: Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Tel.: +31 24 3521301; fax: +31 24 3521213.

E-mail addresses: yangyf@psych.ac.cn (Y. Yang), peter.hagoort@mpi.nl (P. Hagoort).

consistently found negative shifts compared to accented conditions, with varying time windows and scalp topographies: an anterior negativity between 100 and 500 ms in English (Johnson, Clifton, Breen, & Morris, 2003); an anterior negativity between 250 and 1500 ms in Japanese (Ito & Garnsey, 2004); a centro-parietal N400 effect between 200 and 600 ms in German (Hruska & Alter, 2004); a sustained central posterior negative deflection lasting about 500 ms in German (Toepel, Pannekamp, & Alter, 2007); broadly distributed negative effects between 150 and 1050 ms for sentence-final words in French (Magne et al., 2005); a negativity in central posterior areas between 300 and 900 ms in Chinese (Li, Hagoort, & Yang, 2008). Whereas another two studies found a positive shift during 100–750 ms over left hemisphere sites in German (Heim & Alter, 2006, 2007). ERP responses for the opposite case (i.e. non-focused information that is accented) were even less consistent: Johnson et al. (2003), Ito and Garnsey (2004), Hruska and Alter (2004) and Toepel et al. (2007) found no specific effects for the unexpected accentuation, while Magne et al. (2005) observed a positive effect between 300 and 600 ms in the frontal area, and Li, Hagoort, et al. (2008) found an N400 effect (200–700 ms) with central posterior distribution. In the two studies by Heim and Alter (2006, 2007), ERP responses to extra accentuation were dependent on the relative position between accentuation and focused information: extra accentuation that was far ahead of focused information elicited a positive shift between 150 and 250 ms, while extra accentuation that was placed at the word next to the focus information evoked a negative shift over central sites between 100 and 550 ms.

The heterogeneity of these results may be partly explained by between-study differences. For instance, these studies differ in languages, types of context (wh-question, short discourse or single sentence with the particle word “even”), secondary tasks (to judge the appropriateness of the prosodic pattern, to evaluate the discourse meaning, or a delayed lexical probe), and the position of the mismatch between accentuation and focused information (sentence initial, medial or final). Moreover, in all these studies there were methodological issues related to baseline differences. Since one sentence contains two words that have opposite accentuation patterns, the words before the critical words influence the ERPs to the critical words differently in different conditions. Most importantly, given the ERP responses to the correspondence between context and accentuation examined in previous studies, it is difficult to distinguish between the perception of IS and its consequence on semantic processing, that is, the depth of semantic processing.

The notion of “depth of processing” was first put forward by Craik and Lockhart’s (1972) in the framework of memory research. Craik and Lockhart (1972) proposed that the memory trace can be described in terms of depth of processing, with deeper analysis leading to a more persistent trace. This depth-of-processing notion should not be confounded with the depth-of-processing concept often used in language comprehension research (see Sanford & Sturt, 2002, for a review; and Sanford & Graesser, 2006, for another excellent discussion of the concept). In this context, depth of processing refers to the notion that, depending on aspects such as communicative intent, or linguistic context, language is processed either more or less elaborately. In Sanford et al.’s terminology, shallow processing involves an incomplete semantic analysis, which leads to an incomplete, or at least underspecified and/or partial representation of the linguistic input, that may or may not be good enough for the task at hand. Although the conditions under which shallow (as opposed to deep) processing occurs is still an open question, an influential view on shallow processing is that it occurs as a “good enough” strategy that people use for purposes of cognitive efficiency (Ferreira, Bailey, & Ferraro, 2002).

One way of demonstrating the existence of shallow processing has been to show that under certain circumstances people often

do not notice semantic anomalies and inconsistencies. This phenomenon has been termed “semantic illusion”. The so-called Moses illusion (Erickson & Matteson, 1981) is a good example of such a semantic illusion, and refers to the fact that people often miss the anomaly in sentences such as “How many animals of each kind did Moses take on the ark?” Most people answer “two”, failing to notice that it was Noah, not Moses who put animals on the ark.

To measure semantic processing online, a well-known ERP component, the N400, might be a useful tool. The N400 is a negative-going shift, peaking around 400 ms after stimulus onset, with a centro-parietal maximum. It is modulated by the semantic congruency within the context, which could be a particular word, a single sentence or a wider discourse (see Kutas, Van Petten, & Kluender, 2006 for a review). Generally, semantically incongruent words elicit larger N400s than congruent words. This is characterized as an N400 effect (Kutas & Federmeier, 2000). N400 effects are sensitive to semantic processing both in visual and in auditory contexts (Hagoort & Brown, 2000). Moreover, the amplitude of the N400 has been shown to be sensitive to levels of processing (Chwilla, Brown, & Hagoort, 1995).

Since it is a well-established empirical finding that semantic anomalies and/or inconsistencies increase the amplitude of the N400 (the so-called N400 effect), a logical step has been to study semantic illusions using an N400-like experimental paradigm (see, for example, Wang, Hagoort, & Yang, 2009). The rationale here is that shallow processing will result in a decreased sensitivity of the language recipient towards semantic anomalies, which in turn results in a smaller N400 effect. Thus, the size of the N400 effect can be used as a tool to study the effects of various linguistic contexts on the depth of semantic processing.

Recently, we investigated how IS, as realized by a wh-question context, influenced the N400 effect while subjects were reading the answers (Wang et al., 2009). Under two types of wh-question (*What kind of vegetable did Ming buy to cook today?/Who bought the vegetables to cook today?*), the critical word (words underlined) in the answer sentence (*Ming bought eggplant/beef to cook today.*) was either in focus (what-question) or in non-focus (who-question) position, and was semantically congruent (*eggplant*) or incongruent (*beef*). We found a significantly larger N400 effect when critical words were in focus. These results suggest that the N400 effect can be effectively used as a marker for depth of processing during IS manipulations.

As indicated above, prosody is another important parameter for marking IS. Prosodic marking of IS seems to be of a different nature than contextual marking. Whereas context can be considered to have a top-down influence on IS marking, prosodic marking rather relies on bottom-up processes, i.e. the acoustic input features that instantiate accentuation. This difference in IS marking could imply that prosody and context have a relatively independent influence on IS, and related depth of processing. However, in languages such as English and Dutch, in general there is a correspondence between the two parameters context and prosody, such that focus information is accented, and non-focus information is unaccented. This, in contrast, would suggest that the two parameters heavily interact. Therefore, it is important to establish how the two parameters contribute to building IS. Do they exert influence on IS independently, or rather in interaction with each other? The present study addresses these questions in more detail.

In the present study, we auditorily present question-answer pairs to participants. As in the previous study, critical words in the answer sentences were either semantically congruous or incongruous. In addition, critical words were either accented or unaccented, and either in focus or out of focus position. We thus created semantic congruity effects in four combinations: focus/accented (F+A+), focus/unaccented (F+A–), non-focus/accented (F–A+), non-focus/unaccented (F–A–). These manipulations on the critical

Table 1
An example of the eight conditions for one experimental item set.

(a) F+A+S+/S- (what-question, patient accented, semantically congruent/incongruent)	(a) Wat voor soort groente heeft moeder vandaag voor het avondeten gekocht? Vandaag heeft moeder <u>AUBERGINE/RUNDEVLEES</u> voor het avondeten gekocht (What kind of vegetable did mum buy for dinner today? Today mum bought <u>EGGPLANT/BEEF</u> for dinner)
(b) F+A-S+/S- (what-question, agent accented, semantically congruent/incongruent)	(b) Wat voor soort groente heeft moeder vandaag voor het avondeten gekocht? Vandaag heeft <u>MOEDER</u> <u>aubergine/rundvlees</u> voor het avondeten gekocht (What kind of vegetable did mum buy for dinner today? Today MUM bought <u>eggplant/beef</u> for dinner)
(c) F-A+S+/S- (who-question, patient accented, semantically congruent/incongruent)	(c) Wie heeft vandaag de groente voor het avondeten gekocht? Vandaag heeft moeder <u>AUBERGINE/RUNDEVLEES</u> voor het avondeten gekocht (Who bought the vegetable for dinner today? Today mum bought <u>EGGPLANT/BEEF</u> for dinner)
(d) F-A-S+/S- (who-question, agent accented, semantically congruent/incongruent)	(d) Wie heeft vandaag de groente voor het avondeten gekocht? Vandaag heeft <u>MOEDER</u> <u>aubergine/rundvlees</u> voor het avondeten gekocht (Who bought the vegetable for dinner today? Today MUM bought <u>eggplant/beef</u> for dinner)

Note: The examples were originally in Dutch. Literal translations in English are given between brackets. The critical words are underlined, and the accented words are in capitals. F+A+ stands for focus/accented conditions, F+A- stands for focus/unaccented condition, F-A+ stands for non-focus/accented condition, F-A- stands for non-focus/unaccented condition, S+ stands for semantically congruent condition, and S- stands for semantically incongruent condition.

words strictly controlled the interference from the surrounding words, as the semantically congruent and incongruent words were in identical contexts. By comparing the N400 effects obtained in each condition independently, it would give us insight into how the depth of semantic processing is influenced by the different parameters that determine IS. Based on the hypothesis that focused information receives more attention, we expect the largest N400 effect when the focused information is accented (F+A+), while smaller N400 effects are expected to be elicited when there is a mismatch between context and accentuation (F+A-, F-A+), or when there is no focus marking to the new information (F-A-).

2. Methods

2.1. Participants

Thirty-two healthy native speakers of Dutch (16 females, mean age 20, range 18–30 years) were paid to participate in the experiment. They were all right handed, with no dyslexia nor history of hearing problem, nor neurological abnormalities. A consent form according to the Declaration of Helsinki was signed before they started the experiment.

2.2. Stimuli

Question–answer pairs served as the experimental materials. Three different factors were independently manipulated: context, semantic congruency, and (prosodic) accentuation. Context was manipulated through a wh-question, which established a context that projected a focus position in the answer sentence. Each answer contained two constituents, an agent and a patient (see Table 1), one of which was placed in focus position in the answer sentence by the question context. We defined the patient as the critical word (CW). Therefore, in the context of the what-question the CW was focus (F+), and in the context of the who-question the CW was non-focus (F-). Semantic congruency of the CW in the answer was manipulated using a noun that was either semantically congruent (S+) or incongruent (S-) relative to the question context. Note, however, that the local answer sentences were all semantically congruent. It is only in relation with the preceding question that semantic incongruency was established. Finally, accentuation was manipulated through the prosody of the answer sentence. Either the CW (the patient) was accented (A+) or the agent was accented (in which case the CW was not accented, A-). Note that generally, speakers accentuate those words that are in focus position. This means that, in a canonical accentuation pattern, the patient should be accented in the what-question context, and the agent should be accented in the who-question context. Therefore, the F+A+, and the F-A- question–answer pairs (examples a and d in Table 1) have an overall consistent accentuation pattern, whereas the F+A- and the F-A+ pairs (examples b and c in Table 1) have an overall inconsistent accentuation pattern.

Taken together, for the CWs, a full factorial design was created with a combination of three factors: context (F+: focus, F-: non-focus), semantic congruency (S+: congruent, S-: incongruent), and accentuation (A+: accented, A-: unaccented). Therefore, eight conditions were formed for each item set. The semantically congruent and incongruent words were matched on average frequency per million

(mean \pm SD = 822 \pm 1473; 819 \pm 1544, respectively) based on the Dutch CELEX corpus (Baayen, Piepenbrock, & van Rijn, 1993). There was no difference in frequency ($t_{(319)} = .15, p = .88$).

The eight conditions of the 320 item sets were distributed across eight experimental lists according to a Latin square procedure, with each list containing equal number of items (40 items) per condition. In this way we made sure that no single participant listened to a question–answer pair from the same item set more than once, and that each condition of each question–answer pair was presented across the eight different experimental lists. In addition to the experimental items, 160 fillers were constructed in order to make the experimental materials non-transparent for the participants. A 120 fillers were completely correct wh-question answer pairs (including when-, where- and why-questions), 20 fillers contained a semantic anomaly in the answer sentence at positions other than the patient positions, and 20 pairs had inconsistent accentuation at positions other than the patient or agent positions.

Consequently, there were 480 items in each experimental list (320 experimental items, 160 filler items). The eight lists were equally distributed across the 32 subjects. Each list was thus presented to four subjects (two females and two males).

2.3. Acoustic analyses

To make the dialogues as natural as possible, the questions and answers were spoken by a male and a female, respectively, both of whom were native speakers of Dutch. Their dialogues were recorded at 44.1 kHz sampling rate and 16-bit resolution in a soundproof recording room. We used Praat 4.0 (Boersma & Weenink, 2002) to determine the acoustic onsets and offsets of syllables, words and sentences, as well as for further analysis of the duration, intensity and fundamental frequency pattern (F0) of the agent and patient in the answer sentences. The overall sound intensity level was normalized to suppress loudness differences among sentences. To ensure that our female speaker correctly accentuated the relevant words in answer sentences, we performed five one-way ANOVAs for the agent and patient, respectively, with accentuation (accented versus unaccented) as an independent variable, and with word duration, intensity and F0 (mean values, standard deviation and root mean square) as dependent variables. Table 2 presents the results for the acoustic measures and statistical analysis of the agent and patient. The results indicate that the acoustic properties of A+ words differed significantly from A- words. The A+ words had longer durations, higher intensity, higher mean F0, larger F0 standard deviations, and larger F0 rms. Therefore, the acoustic measurements confirmed that the answer sentences were spoken in the intended accentuation pattern. Finally, the average duration of the questions and answers were 3118 ms and 2755 ms, respectively.

2.4. Procedure

Subjects were comfortably seated in a sound-attenuating room, facing a computer screen at approximately 80 cm distance. All materials were presented through loudspeakers at a comfortable listening level. A trial started with a 300 ms warning sound, and 700 ms later the question was presented. After a silence of 1000 ms, the corresponding answer was presented. A fixation cross appeared on the center of the screen, 500 ms before the end of the question and remained present until 1000 ms after the end of the answer. It was replaced by three scores presented for 2000 ms. Then the next trial began. The participants were required not to move or blink during the fixation cross period, while they were encouraged to blink during the appearance of the scores. The question–answer-pairs were presented in a pseudorandom

Table 2
Acoustic measurements of the agent and patient words in the target answer sentences. The patient is the critical word (CW) in the ERP measurements.

	Agent			Patient (CW)		
	A+	A–	F(1,319)	A+	A–	F(1,319)
Duration (ms)	442.73 (113.70)	395.75 (118.80)	682.43***	493.00 (129.68)	454.99 (127.58)	537.375***
Intensity (dB)	72.25 (2.35)	69.19 (2.43)	1725.11***	72.52 (2.15)	65.89 (2.50)	3898.76***
F0 mean (Hz)	266.08 (22.97)	211.80 (12.27)	1993.47***	252.78 (22.36)	187.36 (9.51)	3165.14***
F0 std	47.72 (11.26)	12.62 (6.32)	2786.90***	44.24 (11.64)	12.77 (9.75)	1391.26***
F0 rms	.11 (.03)	.06 (.02)	1367.34***	.09 (.02)	.04 (.01)	2646.38***

Note: Means and standard deviations (in brackets) are presented. Mean stands for the mean value, std stands for the standard deviation, and rms stands for the root mean square. A+ represents accented condition while A– represents unaccented condition.

*** Significance at .001 level.

order, such that no more than three items of the same condition were presented in succession. A new pseudorandom order was created for each participant.

The 480 items in one list were divided into 24 blocks (20 trials/block), with each block lasting about 4 min. To avoid an exceedingly long recording session, the experiment was separated into two sessions (12 blocks per session). Each session took about 2 h, including subject preparation, instructions and a short practice block consisting of 15 items. The subjects finished these two sessions at two different times, with at least 3 days intervening.

2.5. EEG recordings

The EEG was recorded in an electromagnetically shielded cabin, with 60 surface active electrodes (Acticap, Brain Products, Herrsching, Germany) placed in an equidistant montage. The left mastoid electrode served as the reference, and a forehead electrode served as the ground. A supra- to suborbital bipolar montage was used to monitor vertical eye movements, while a right to left canthal bipolar montage was used to monitor horizontal eye movements. All electrode impedances were kept below 20 k Ω during the experiment (which is well below what is recommended for active electrodes). EEG data were digitized at a rate of 500 Hz with a 100 Hz high cut-off filter and a 10 s time constant.

2.6. Data analysis

Brain Vision Analyzer software 1.05 (Brain Products) was used to preprocess the raw EEG data. The EEG data were re-referenced off-line to the average of both mastoids, and lowpass filtered at 30 Hz (48 dB/oct slope). Next, the data were segmented from 200 ms before to 1200 ms after the onset of the critical words, with baseline correction from –200 to 0 ms preceding word onset. After this step, a semi-automatic artifact rejection procedure was applied. On average 6% of all trials were rejected, with rejections being equally distributed across the eight conditions ($F < 1$). Finally, trials were averaged in each condition for each subject, and this average was used for further statistical analysis.

The statistical significance of the difference between two conditions was evaluated by a cluster-based random permutation approach (see Maris & Oostenveld, 2007 for details on the method), which was implemented in the Matlab toolbox Fieldtrip (<http://fieldtrip.fcdonders.nl/>). This approach elegantly handles the multiple comparisons problem (there were 600 time points * 59 channels = 35,400 comparisons in total). The approach naturally takes care of interactions between time points and electrodes by identifying clusters of significant differences between conditions in the time and space dimensions, and effectively controls the Type-1 error rate in the case of multiple comparisons. Here is a brief description of the procedure (for more details, see Maris & Oostenveld, 2007).

First, for every data point (time-electrode point) a simple dependent-samples *t*-test is performed (giving uncorrected *p*-values). All data points that do not exceed a pre-set significance level (here 5%) are zeroed. Clusters of adjacent non-zero data points are computed, and for each cluster a cluster-level test statistic is calculated by taking the sum of all the individual *t*-statistics within that cluster. Next, a null-distribution is created as follows. Subject averages are randomly assigned to one of the two conditions several times (1000 times here), and for each of these randomizations, cluster-level statistics are computed. For each randomization, the largest cluster-level statistic enters into the null distribution. Finally, the actually observed cluster-level test statistics are compared against the null distribution, and clusters falling in the highest or lowest 2.5th percentile are considered significant.

This procedure only allows for pair-wise comparisons. Therefore, a set of four pair-wise comparisons was performed to test the semantic congruity effects in four conditions: F+A+ (F+A+S– versus F+A+S+), F+A– (F+A–S– versus F+A–S+), F–A+ (F–A+S– versus F–A+S+) and F–A– (F–A–S– versus F–A–S+). All 59 electrodes (excluding the right mastoid electrode), and 600 time points (corresponding to an interval from CW onset to 1200 ms after CW onset, at 500 Hz sampling rate) were entered into the analysis for the grand averaged data.

To further evaluate the modulations of semantic congruity effects by context and accentuation, we took the difference waveforms (DWs) of individual subject for each electrode, which were obtained by subtracting the ERPs elicited by semantically congruent condition from those of the incongruent condition in each IS condition

(F+A+, F+A–, F–A+, F–A–), as the input of the cluster-based randomization statistics. Based on the statistical results of the semantic congruity effects, 59 electrodes and 500 time points (corresponding to the interval from 200 to 1200 ms after CW onset) were included in the tests. For the statistical analysis of the DWs, we examined the main effects of context and accentuation, as well as their interaction. The main effect of context was obtained by comparing the DW of F+ (the averaged DWs of F+A+ and F+A–) with F– (the averaged DWs of F–A+ and F–A–); similarly, the main effect of accentuation was acquired by comparing the DW of A+ (the averaged DWs of F+A+ and F–A+) with A– (the averaged DWs of F+A– and F–A–). Then the interaction between context and accentuation was tested by comparing two subtractions: (F+A–) – (F+A+) versus (F–A–) – (F–A+). If the interaction was significant, further simple effect analysis were conducted.

3. Results

The nonparametric statistics for the entire set of 32 subjects revealed significantly larger negativities for incongruent words relative to congruent words in all four conditions (F+A+, F+A–, F–A+ and F–A–). Fig. 1 displays the semantic congruity effects at a representative electrode (corresponding to location Cz in a standard 10–20 set-up) for four conditions. These effects were mainly distributed in the central-posterior area, with the significant time window between 254 and 954 ms for F+A+ ($p < .001$), between 380 and 752 ms for F+A– ($p = .03$), between 398 and 734 ms ($p < .001$) for F–A+, and between 346 and 760 ms ($p = .01$) for F–A– (see Table 3). Given their topographical and latency characteristics, we classified these effects as N400 effects. Note that the N400 effects showed a broader morphology than the classical N400 effects obtained in visual presentation studies (Kutas & Federmeier, 2000). This might be related to the auditory presentation mode, where the exact duration of auditory stimuli varies across trials (see Hagoort & Brown, 2000; Li, Hagoort, et al., 2008 for similar broadly distributed N400s).

Visual inspection of the N400 effects shows that the F+A+ condition elicited the largest N400 effect, which was supported by the statistical analysis on the condition difference. Testing the difference in effect sizes revealed that F+ elicited a larger N400 effect than F– in the time windows of 664–888 ms ($p < .001$); A+ elicited a larger N400 effect than A– between 646 and 882 ms ($p = .05$). Moreover, the interaction between context and accentuation was significant in the time window between 666 and 890 ms ($p = .02$). Simple effect tests revealed different accentuation effects for F+ and F– condition. A larger N400 effect was found for F+A+ than F+A– between 646 and 882 ms ($p = .05$), while no significant difference was found in the N400 effect between F–A+ and F–A– ($p = .28$). Similarly, there was a different context effects for the A+ and the A– condition. We found a larger N400 effect for F+A+ than F–A+ between 664 and 890 ms ($p = .02$), while F+A– and F–A– did not show any significant N400 difference ($p = .45$).

Next, in an exploratory analysis we split up the data by subjects' sex, and found substantial differences between males and females (see Figs. 2 and 3). We grouped the subjects into females ($N = 16$) and males ($N = 16$) and tested the N400 effects modulated by context and accentuation in each of these groups. For females, the incongruent words elicited a larger N400 than congruent words

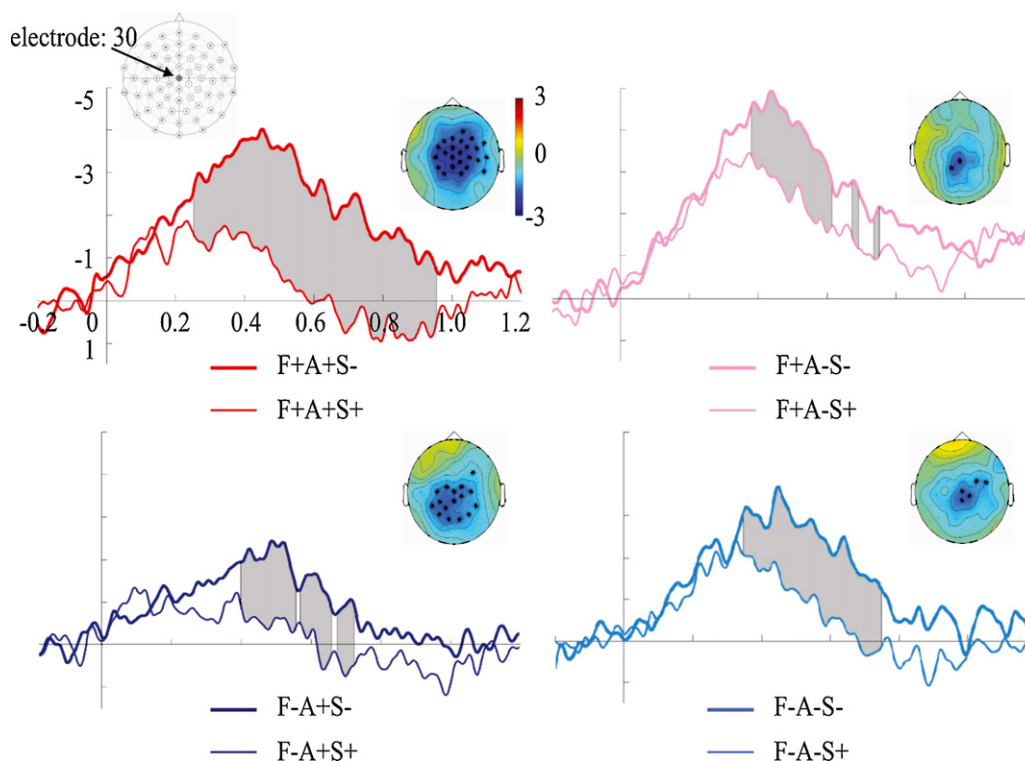


Fig. 1. The N400 effects elicited by incongruent words relative to congruent words for all the subjects ($N=32$). The onset of the critical words is at zero. In each condition, the thick line represents the waveforms elicited by incongruent words, while the thin line represents the waveforms elicited by congruent words. The red lines are for the F+A+ condition (focus/accented), the pink lines are for the F+A– condition (focus/unaccented), the dark blue lines are for the F–A+ condition (non-focus/accented), and the light blue lines are for the F–A– condition (non-focus/unaccented). The time intervals with significant effects are painted grey, and the topographies were plotted in the time interval of the electrode that showed the largest effect, with the electrodes that showed significant effects over 85% of the selected time intervals marked by *. Note that negative is plotted up. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between 342 and 898 ms for F+A+ ($p=.01$), between 416 and 684 ms for F+A– ($p=.04$), between 386 and 662 ms for F–A+ ($p=.003$), and between 376 and 740 ms for F–A– ($p=.01$) (see Table 3). Although the visual inspection showed the largest N400 effect for F+A+, the statistical analyses performed on the effect sizes of four conditions revealed no significant differences. These results suggested that for females, the N400 effects did not differ as a function of context or accentuation. The males, however, showed a different pattern of results (see Fig. 3). The semantically incongruent words elicited a significantly larger N400 than the congruent words only for F+A+ condition between 226 and 804 ms ($p=.004$), whereas the other three conditions (F+A–, F–A+ and F–A–) revealed no significant N400 effects (all p -values $>.10$, see Table 3). Testing for condition differences indicated that the F+A+ condition elicited a larger N400 effect than the other conditions.

When explicitly testing sex (female, male) in a mixed-model approach (with context and accentuation as within-subject factors, and sex as a between-subject factor), we did not find any significant interaction between sex and the other factors. Nevertheless, the

different patterns of N400 effects for females and males, and the analyses on the separate groups hint at possible sex difference in the semantic processing.

For visual illustration of the N400 effects, three sets of DWs in each of the four conditions are shown in Fig. 4, with Fig. 4(a) for grand average, Fig. 4(b) for females and Fig. 4(c) for males.

4. Discussion

The purpose of the current study was to examine the relative contributions of context and accentuation on the depth of semantic processing. Depth of semantic processing was evaluated by measuring N400 effects elicited by semantically incongruent words relative to congruent words in question–answer pairs. We found the largest N400 effect when the focused information was accented (F+A+), while smaller N400 effects occurred in the other three conditions (F+A–, F–A+, F–A–). A possible sex difference in the depth of semantic processing was supported by separate analyses on females and males.

Table 3

The statistical results of the semantic congruity effects in the four conditions.

	Grand average		Females		Males	
	Time window	p	Time window	p	Time window	p
F+A+	254–954 ms	$p < .001$	342–898 ms	$p = .01$	226–804 ms	$p = .004$
F+A–	380–752 ms	$p = .03$	416–684 ms	$p = .04$	NA	n.s.
F–A+	398–734 ms	$p < .001$	386–662 ms	$p = .003$	NA	n.s.
F–A–	346–760 ms	$p = .01$	376–740 ms	$p = .01$	NA	n.s.

Note: n.s.: not significant; NA: not applicable. F+A+ stands for focus/accented condition, F+A– stands for focus/unaccented condition, F–A+ stands for non-focus/accented condition, and F–A– stands for non-focus/unaccented condition.

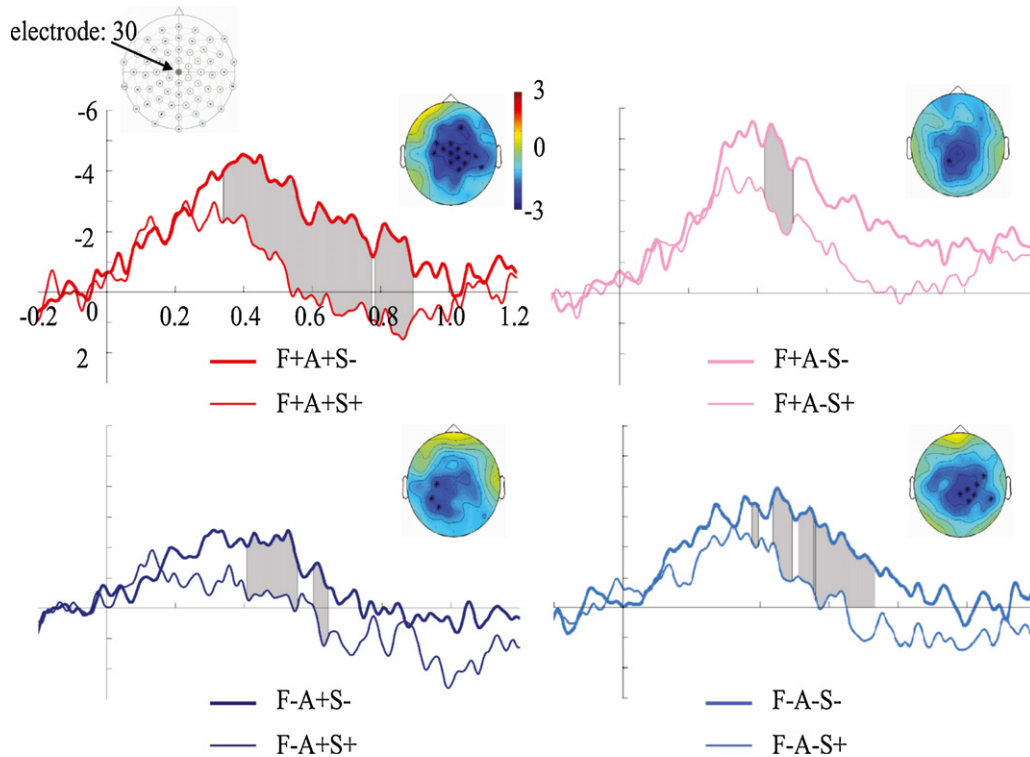


Fig. 2. The N400 effects elicited by incongruent words relative to congruent words for females ($N = 16$). The onset of the critical words is at zero. For detailed label information, see the legend of Fig. 1.

4.1. More elaborative semantic processing in F+A+ condition than the other conditions

Semantic incongruency between question and answer produced the N400 effect in the F+A+ condition between 254 and 954 ms

at central electrodes, whereas smaller N400 effects were found in the other three conditions (F+A-, F-A+, F-A-). The divergence between F+A+ and the other three conditions is in line with the findings of behavioral studies. For instance, it was found that speech comprehension is facilitated in sentences with proper accentuation

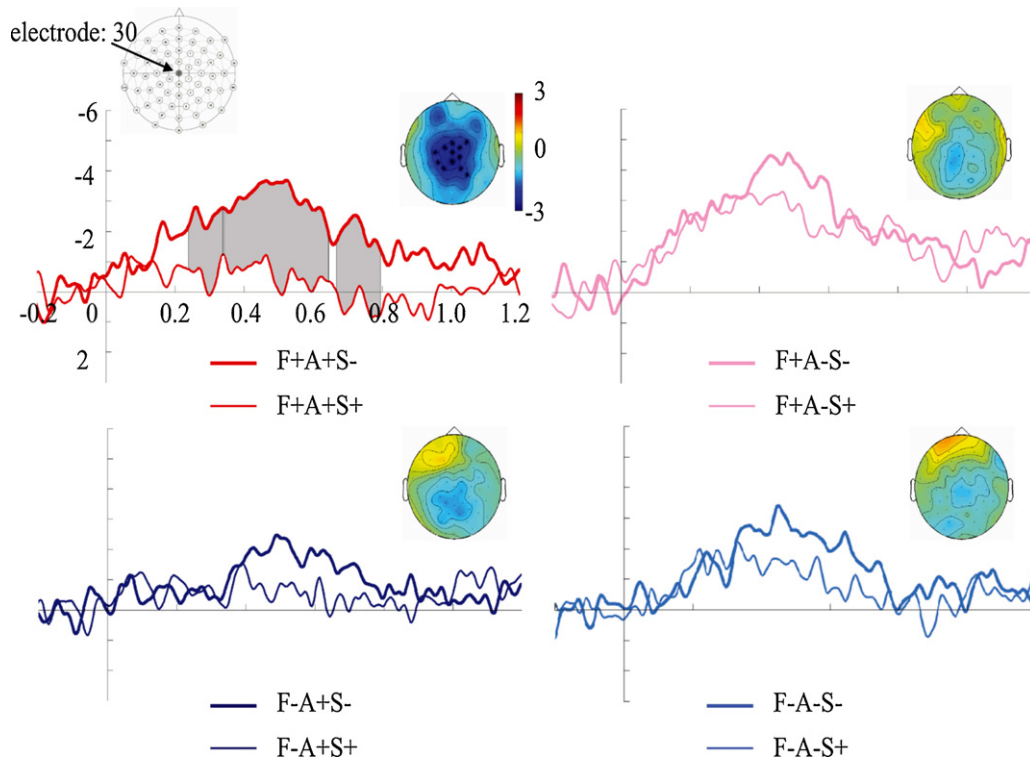


Fig. 3. The N400 effects elicited by incongruent words relative to congruent words for males ($N = 16$). The onset of the critical words is at zero. For detailed label information, see the legend of Fig. 1. For the conditions that show no significant effect, the topographies were plotted between 0.3 and 0.6 s.

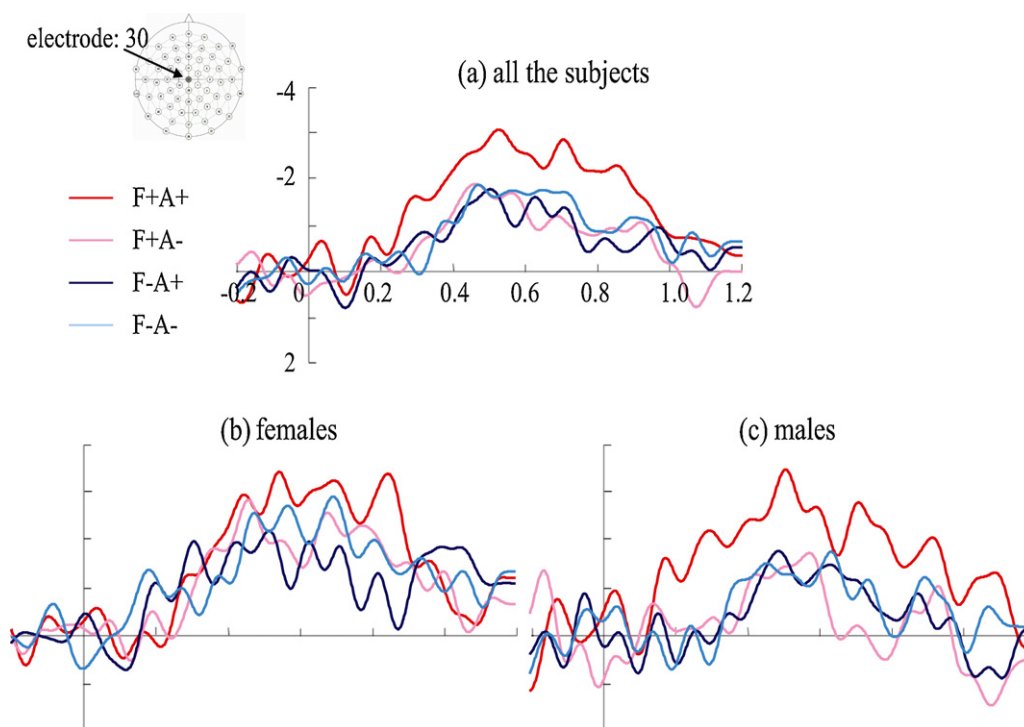


Fig. 4. The difference waveforms between incongruent and congruent words for all the subjects (a), females (b), and males (c), respectively. The onset of the critical words is at zero. The red lines are for the F+A+ condition (focus/accented), the pink lines are for the F+A- condition (focus/unaccented), the dark blue lines are for the F-A+ condition (non-focus/accented), and the light blue lines are for the F-A- condition (non-focus/unaccented). The waveforms were 10 Hz low pass filtered for illustration purposes only. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Bock & Mazzella, 1983; Dahan et al., 2002; Terken & Nootboom, 1987). In the current experiment, the question context generated a strong expectation with regard to when new information would become available. The corresponding accentuation further recruited sufficient attentional resources for adequate semantic processing at the focus position (Birch & Rayner, 1997; Cutler & Fodor, 1979; Ward & Sturt, 2007). Therefore, an elaborate processing took place for the F+A+ condition, where the critical words in the semantically incongruent versus congruent conditions resulted in an N400 effect.

In contrast, smaller N400 effects were found for the other three conditions (F+A-, F-A+, F-A-). This suggests that in these conditions the semantic information was less elaborately processed (Wang et al., 2009). In the following, we will discuss the more superficial semantic processing in these three conditions.

Relative to the F+A+ condition, the missing of accentuation in the F+A- condition and the lack of context marking in the F-A+ condition induced smaller N400 effects around 650–890 ms. The similar temporal pattern for F+A- and F-A+ enables us to take the reduced N400 effects as a manifestation of less elaborate semantic processing due to the mismatch between the context and the accentuation pattern. In both the F-A+ and the F+A- conditions, the wh-question contexts led to different expectations about how IS marking of the answer would be realized. The missing accentuation on the focused information (F+A-) and extra accentuation on the non-focused information (F-A+) reduced the depth of semantic processing (Cutler & Fodor, 1979; Dahan et al., 2002; Hruska & Alter, 2004; Li, Hagoort, et al., 2008; Terken & Nootboom, 1987). The results (especially the significant interaction between context and accentuation) suggest that context and accentuation are not independent processing streams, but instead actively interact with each other.

Although there is no conflict between context and accentuation in the F-A- condition, it also produced a smaller N400 effect than the F+A+ condition. This might reflect the reduced integration

when the critical word information was neither in focus position nor in accented position. When the critical word information was in non-focus position, the who-question contexts generated a strong expectation that the agent is the focus, with less attention allocated to the non-focus position. Furthermore, the accentuation in the agent position recruited more processing resources towards the agent. Therefore, the incoming, unaccented information (patient) that appeared in non-focus position might have received fewer attentional resources, with less detailed semantic processing as a result.

Overall, these results confirm the findings of other ERP studies (Heim & Alter, 2006, 2007; Hruska & Alter, 2004; Ito & Garnsey, 2004; Johnson et al., 2003; Li, Yang, & Hagoort, 2008; Li, Hagoort, et al., 2008; Magne et al., 2005; Toepel et al., 2007) by showing how context and accentuation influence the depth of semantic processing. In ERP studies on the role of IS, ERP responses with distinct time courses were found under the influence of context and accentuation. The effect starts as early as 100 ms (Johnson et al., 2003) and ends as late as 1500 ms (Ito & Garnsey, 2004). In the current study, we focused on the consequences of IS on semantic processing. We found that the contextually determined focus and the accentuation pattern need to converge in order to attain deep semantic processing. Otherwise, extended semantic processing seems to be reduced due to the interference between the anticipatory top-down expectation and the unexpected bottom-up input (F+A- and F-A+), or the lack of a trigger for the recruitment of attentional resources (F-A-). The similarly reduced N400 effects for the F+A-, F-A+ and F-A- conditions relative to the F+A+ condition suggest that top-down (context) and bottom-up (accentuation) markers of IS modulate the attentional resources in a similar manner.

4.2. Sex differences in the depth of semantic processing

Although it was not anticipated, we found that females and males showed different patterns of the N400 effects under the four

conditions (albeit that the effect was not strong enough to reach significance in a mixed-model analysis). Females produced significant N400 effects in all four conditions (F+A+, F+A–, F–A+, F–A–), whereas males only generated an N400 effect in the F+A+ condition, with no reliable N400 effects occurring in the other three conditions (F+A–, F–A+, F–A–). This sex difference is consistent with some studies which suggest that females produce larger N400 effects than males in response to semantically unrelated words or sentences (Daltrozzo, Wioland, Kotchoubey, 2007; Wirth et al., 2007), and that females are better at detecting conflict between semantic and pragmatic information in a linguistic context (Van den Brink et al., 2010). The idea that females are more likely to engage in deeper semantic processing than males is also in line with more generally observed gender differences in cognitive style. Previous cognitive studies have found that females process information in an exhaustive way, and that they rely on all available information before rendering judgment. In contrast, male information processing is usually partial and incomplete, relying on a subset of highly available and salient cues instead of detailed message elaboration (Guillem & Mograss, 2005; Meyers-Levy & Sternthal, 1991; Skrandies, Reik, & Kunze, 1999). We conclude that sex differences in semantic processing clearly deserve further investigation.

Acknowledgements

This study was supported by the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Chinese Academy of Sciences (CAS). The authors thank Jacqueline de Nooijer for the construction of materials. We also thank Petra van Alphen and Joost Rommers for recording materials.

References

- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia: University of Pennsylvania.
- Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, 25(5), 653–660.
- Bock, J. K., & Mazzella, J. R. (1983). International marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, 11(1), 64–76.
- Boersma, P., & Weenink, D. (2002). *Praat 4.0: A system for doing phonetics with the computer [computer software]*. Amsterdam: Universiteit van Amsterdam.
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32(3), 274–285.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
- Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, 7, 49–59.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47(2), 292–314.
- Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2007). Sex differences in two event-related potentials components related to semantic priming. *Archives of Sexual Behavior*, 36(4), 555–568.
- Erickson, T. A., & Matteson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540–552.
- Ferreira, F., Bailey, Karl, G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15.
- Guillem, F., & Mograss, M. (2005). Gender differences in memory processing: Evidence from event-related potentials to faces. *Brain and Cognition*, 57, 84–92.
- Günther, C., Maienborn, C., & Schopp, A. (1999). The processing of information structure. In P. Bosch, & R. A. van der Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 18–43). Cambridge University Press.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, 38(11), 1518–1530.
- Heim, S., & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis*, 66(1), 55–68.
- Heim, S., & Alter, K. (2007). Focus on focus: The brain's electrophysiological responses to focus particles and accents in German. In A. späth (Ed.), *Language context & cognition: Interfaces and interface conditions*. Berlin: De Gruyter.
- Hruska, C., & Alter, K. (2004). Prosody in dialogues and single sentences: How prosody can influence speech perception. In A. Steube (Ed.), *Information structure: Theoretical and empirical aspects* (pp. 221–226). Berlin: Walter de Gruyter.
- Ito, K., & Garnsey, S. M. (2004). Brain responses to focus-related prosodic mismatch in Japanese. *Proceedings of Speech Prosody*, 609–612.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York: Oxford University Press.
- Johnson, S. M., Breen, M., Clifton, C. J. R., & Morris, J. (2003). ERP investigation of prosodic and semantic focus. In *Poster presented at cognitive neuroscience* New York City.
- Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. A. Gernsbacher, & M. Traxler (Eds.), *Handbook of psycholinguistics* (pp. 659–724). New York: Elsevier Press.
- Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognition Science*, 4(12), 436–470.
- Li, X., Yang, Y., & Hagoort, P. (2008). Pitch accent and lexical tone processing in Chinese discourse comprehension: An ERP study. *Brain Research*, 1222, 192–200.
- Li, X., Hagoort, P., & Yang, Y. (2008). Event-related potential evidence on the influence of accentuation in spoken discourse comprehension in Chinese. *Journal of Cognitive Neuroscience*, 20(5), 906–915.
- Magne, C., Astesano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of “pop-out” words in spoken French dialogues. *Journal of Cognitive Neuroscience*, 17(5), 740–756.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 5(1), 177–190.
- Meyers-Levy, J., & Sternthal, B. (1991). Gender differences in the use of message cues and judgments. *Journal of Marketing Research*, 28, 84–96.
- Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes*, 42(2), 99–108.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognition Science*, 382(9), 382–386.
- Skrandies, W., Reik, P., & Kunze, C. (1999). Topography of evoked brain activity during mental arithmetic and language tasks: Sex differences. *Neuropsychologia*, 37, 421–430.
- Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, 2(3), 145–163.
- Toepel, U., Pannekamp, A., & Alter, K. (2007). Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioural and Brain Functions*, 3, 53.
- Van den Brink, D. L., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., et al. (2010). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, doi:10.1093/scan/nsq094
- Wang, L., Hagoort, P., & Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain Research*, 1282(28), 50–56.
- Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition*, 35, 73–86.
- Wirth, M., Horn, H., Koenig, T., Stein, M., Federspiel, A., Meier, B., et al. (2007). Sex differences in semantic processing: Event-related brain potentials distinguish between lower and higher order semantic analysis during word reading. *Cerebral Cortex*, 17(9), 1987–1997.