PSYCHOPHYSIOLOGY SPR SOCIETY FOR PSYCHOPHYSIOLOGICAL RESEARCH

Who are you talking to? The role of addressee identity in utterance comprehension

| Abstract: | Experimental evidence suggests that speaker and addressee quickly adapt to each other from the earliest moments of sentence processing, and that interlocutor-related information is rapidly integrated with other sources of non-pragmatic information (e.g, semantic, morphosyntactic, etc.). These findings have been taken as support for one-step models of sentence comprehension. The results from the present ERP study challenge this theoretical framework providing a case where discourse level information is integrated only at a late stage of processing, when morphosyntactic analysis has been already initiated. We considered the case of Basque allocutive agreement, where information about addressee gender is encoded in verbal inflection. Two different types of Basque grammatical violations were presented together with the corresponding control conditions: one could be detected based on a morphosyntactic mismatch (person agreement violation), while the other could be detected only if the addressee's gender was considered (allocutive violation). Morphosyntactic violations elicited greater N400 effects followed by P600 effects, while allocutive violations elicited only P600 effects. These results provide new constraints to one-step accounts as they represent a case where speakers do not immediately adjust to the addressee's perspective. We propose that the relevance of discourse-level information might be a crucial variable to reconcile the dichotomy between one- and two-step models. |

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The present electrophysiological findings advance our knowledge in multiple theoretical domains: electrophysiology, social cognitive psychology, and psycholinguistics. From an electrophysiological point of view, the present results advance our understanding of the relation between cognitive processes and neural responses, proposing a functional distinction between syntax-related and pragmatics-related P600 effects. The findings also provide new theoretical constraints for cognitive models of referential communication, supporting the late integration of common ground knowledge. Finally, within the psycholinguistic domain, the results contribute reconciling and overcoming the apparent dichotomy between one-step and two-step models of parsing.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Running head: ERP FINDINGS ON THE ROLE OF ADDRESSEE IDENTITY

Who are you talking to? The role of addressee identity in utterance comprehension

Sendy Caffarra[a], Max Wolpert[b,c], Dana Scarinci[a], Simona Mancini[a]

[a] BCBL, Basque Center on Cognition, Brain and Language, 69 Mikeletegi, San Sebastian,

Spain 20009

[b] McGill University Integrated Program in Neuroscience, 3801 University Street Montreal,

Quebec H3A 2B4

[c] Centre for Research on Brain, Language and Music, 3640 de la Montagne, Montreal QC

H3G 2A8

Corresponding author:

Sendy Caffarra

BCBL, Basque Center on Cognition, Brain and Language, Donostia, Spain

Paseo Mikeletegi 69,

20009 Donostia-San Sebastian,

Spain.

tel.: +34 943 309 300

email: s.caffarra@bcbl.eu

1

Abstract

Experimental evidence suggests that speaker and addressee quickly adapt to each other from the

earliest moments of sentence processing, and that interlocutor-related information is rapidly integrated

with other sources of non-pragmatic information (e.g, semantic, morphosyntactic, etc.). These

findings have been taken as support for one-step models of sentence comprehension. The results from

the present ERP study challenge this theoretical framework providing a case where discourse level

information is integrated only at a late stage of processing, when morphosyntactic analysis has been

already initiated. We considered the case of Basque allocutive agreement, where information about

addressee gender is encoded in verbal inflection. Two different types of Basque grammatical

violations were presented together with the corresponding control conditions: one could be detected

based on a morphosyntactic mismatch (person agreement violation), while the other could be detected

only if the addressee's gender was considered (allocutive violation). Morphosyntactic violations

elicited greater N400 effects followed by P600 effects, while allocutive violations elicited only P600

effects. These results provide new constraints to one-step accounts as they represent a case where

speakers do not immediately adjust to the addressee's perspective. We propose that the relevance of

discourse-level information might be a crucial variable to reconcile the dichotomy between one- and

two-step models.

Keywords: speech comprehension, addressee identity, ERPs

1 Introduction

Social aspects of the conversational setting are crucial to correctly understand an utterance. For example, what we know about the person a statement is directed to helps us avoid misinterpretations and identify the intended message. The sentence "You're very agile!" might be interpreted literally if it is said to an athlete, but it can be a joke if it is said to someone with a broken leg. Addressee identity can have even stronger consequences for sentence interpretation. The utterance "You're tanned!" is an objective description if directed to someone who sunbathed, but it becomes a racist insult when used to address an African-American politician (Glendinning, 2008). The present study will investigate at which point in time information about addressee identity is taken into account during sentence processing.

In the sentence comprehension literature, different theoretical models have been proposed to describe how and when pragmatic information (such as properties related to the speaker and the addressee of an utterance and relative inferences) is integrated and interacts with other levels of language analysis. Two alternative descriptions of the time course of pragmatic analysis have been proposed.

Two-step models assume that the first stage of sentence comprehension is encapsulated and purely syntactic, meaning that it is uniquely based on the structural properties of the language input which do not interact with other sources of information (e.g., information about the gender of the interlocutor during the comprehension of the utterance "I'm pregnant"). It is only at a later stage of processing where information coming from non-syntactic domains would be integrated and could potentially interact with the initial syntax-based interpretation (Fodor, 1983; Frazier, 1987; Frazier & Clifton, 1996; Friederici, 2002; Forster & Ryder, 1971; cfr. Steinhauer & Drury, 2012). Thus, relevant information about addressee identity is supposed to have a delayed effect with respect to syntactic information.

In contrast, one-step models are highly interactive and claim that different sources of information can be taken into account at an early stage of processing. Any contextual constraint, no matter its source, can be quickly available and processed (Jackendoff, 2002; Marsel-Wilson & Tyler, 1980; McDonald,

3

Pearlmutter & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). Hence, potentially useful information about the individuals involved in a communicative context is integrated early and can potentially constrain subsequent phases of analysis.

Electrophysiological methods, such as event-related potentials (ERPs), have been widely used to provide a fine-grained temporal picture of pragmatic analysis and tease apart these alternative theoretical proposals. Among all ERP studies focused on discourse-level information, speaker identity has received considerable attention.

Lattner and Friederici (2003) conducted an ERP study presenting German spoken sentences whose content was either congruent or not with speaker gender stereotypes ("I like to play soccer" said by a man or a woman). Effects of speaker congruence were observed at a relatively late phase of analysis (600 ms after the relevant stereotypical information was available; e.g., "soccer"), with greater posterior positivities for incongruent as compared to congruent speakers. This P600 effect was interpreted as reflecting late integration processes of speaker-related knowledge, which typically occurred after syntactic and semantic analysis (Friederici, 2002). Although these findings are in line with the two-step model's predictions (treating the P600 component as one of the ERP markers of the second step of analysis), more recent ERP studies using similar paradigms came to opposite conclusions (Bornkessel-Schlesewsky, Krauspenhaar & Schlesewsky, 2013; Van Berkum, van den Brink, Tesink, Kos & Hagoort, 2008). When the content of the utterance was not congruent, or at least implausible, with respect to speakers' characteristics (e.g., gender, social status, age, political relevance) greater posterior negativities were observed as compared to speaker-congruent sentences. Crucially, this effect appeared at an early stage of analysis (200 ms after the information was available) and reached its maximum amplitude around 400 ms after target presentation. This ERP effect did not differ in latency and topography from N400 effects observed in response to sentences that were semantically incongruent regardless of speaker identity (e.g., "you wash your hands with horse"; Van Berkum et al., 2008). These findings have been interpreted as supporting one-step models, showing that what is inferred from the speaker's voice is quickly available and impacts early stages of meaning construction. Additional findings suggest that contextual knowledge can impact

4

early stages of syntactic analysis as well, with early pragmatic effects reported when speaker identity was incongruent with the morphological information of the utterance. Hanulíková and Carreiras (2015) conducted an ERP study on Slovak spoken sentences where information about speaker gender was conveyed by past participle inflections (e.g., *lebo som kradla/kradol slivky*, I have stolen$_F$/stolen$_M$ plums, where the inflection of the verb 'stolen' indicates the gender of the speaking subject). Greater N400 effects (100-400 ms after inflection onset) were observed with incongruent as compared to congruent speaker, demonstrating that speaker identity impacts morphosyntactic processing at the earliest stages.

It is worth noting that the types of experimental manipulations used by the ERP studies described above maximized the potential relevance of speaker characteristics. Most studies used self-referential pronouns (except for Bornkessel-Schlesewsky et al., 2013), defining the speaker as the protagonist of the speech act. The content of the utterance was typically self-oriented and speaker characteristics could be inferred directly from voice variations available at the moment of the speech act. Less attention has been given to other discourse roles that might be less salient, such as the addressee. The addressee is placed at a certain distance from the speaker, which is considered the origin of the deictic field (i.e., the reference in relation to which different referential expressions acquire their context-dependent meaning; Bühler, 1982). In addition, at the moment of the speech act, addressee-related characteristics cannot be inferred from the speaker's voice characteristics.

Cognitive models of referential communication have described the processing of addressee-related knowledge in utterance comprehension and put forward two different proposals that resemble the abovementioned distinction between one-step and two-step models (Brennan, Galati & Kuhlen, 2010, for a review). Two-step proposals claimed that the knowledge shared between the speaker and the addressee (i.e., common ground) does not affect the first stage of utterance interpretation (Horton & Keysar, 1996; Keysar, Barr & Horton, 1998; Pickering & Garrod, 2004). The listeners initially assume an egocentric perspective, which is not sensitive to pragmatic constraints. This early and automatic phase would be followed by a late stage including inferences about interlocutor-related information. The alternative proposal claims that common ground knowledge shapes the early stages

5

of real-time comprehension and production planning (Clark & Carlson, 1981; Galati & Brennan, 2010; Hanna, Tanenhaus & Trueswell, 2003). Eye-tracking studies aimed at teasing apart these positions have typically adopted collaborative tasks (e.g., helping a confederate find an object) where participants were presented with utterances containing referential expressions (e.g., "Find the chocolate sprinkles"). Participants' tasks consisted of visually localizing the correct referent among different objects, some of which were visible only to one interlocutor (privileged ground) and some to both (common ground). A few studies found no early effects of common ground knowledge, suggesting that mutual knowledge might be used at a later stage, and only if a revision of the interpretation was needed (Keysar, Balin & Brauner, 2000; Keysar, Lij & Barr, 2003). The vast majority of eye-tracking studies on this topic, however, showed that early stages of participants' visual search were affected by common ground knowledge, supporting an immediate use of interlocutor-based constraints during reference resolution (Brennan, Galati & Kuhlen, 2010, for a review; Brown-Schmidt, Gunlogson & Tanenhaus, 2008; Nadig & Sedivy, 2002; Hanna et al., 2003; Hanna & Tanenhaus, 2004; Heller, Grodner & Tanenhaus, 2008). Similar to the ERP literature, these early eye-tracking effects were localized within 500 ms after critical stimulus onset. Importantly, this evidence comes from a collaborative setting, where the information about addressee was highly relevant to the participants' interactive task.

The present work advances a step further in the study of addressee identity analysis by studying effects in a non-collaborative dialogic context. Also, interlocutor-based knowledge will be available in the utterance, although not crucial for sentence comprehension.

1.1 The present study

The present study will investigate the temporal dynamics of the analysis of addressee-related information during utterance perception. We will consider the case of Basque, a language spoken in northern Spain and southwestern France. The Basque language is ideal to address the present research question as it can encode addressee-related information in the morphosyntactic structure of the utterance. Specifically, in Basque allocutive agreement, the auxiliary verb inflection must agree in

gender with the sex of the addressee when the addressee is not an argument of the verb (Oyharçabal, 1993; Hualde & Ortiz de Urbina, 2003). Allocutive agreement is mandatory when a familiar register is used (Oyharçabal, 1993), which may be socially restricted to conversations by and between men and older speakers (e.g., Echeverria, 2003, 2010). Note that the use of allocutive (especially in its feminine forms) is in danger of being lost and today only a subset of Basque speakers are still proficient with using allocutive agreement (Haddican, 2003)[1]. Our study specifically focuses on these proficient allocutive users.

To test the impact of addressee identity in sentence processing, simple Basque conversations consisting of two utterances produced by two distinct interlocutors (whose role of speaker and addressee switched at each turn-taking) were created and presented to Basque native listeners (overhearers). In each conversation there was a target utterance containing a Basque allocutive verb that either agreed or disagreed with the addressee's gender. Note that the target utterance was perfectly grammatical in isolation (i.e., without considering the rest of the dialogue), and the incongruence was evident only when both target verb inflection and addressee-related information were considered.

Additional conversations were included where the target utterances could show a morphosyntactic violation (grammatical person disagreement) or the corresponding grammatical construction. Unlike the allocutive violations, these grammatical incongruences were detectable regardless of addressee identity. This control comparison was added to monitor the time course of morphosyntactic analysis when interlocutor's characteristics were not relevant.

Pragmatic and morphosyntactic analysis were monitored millisecond by millisecond using ERPs. Morphosyntactic violations were expected to elicit greater negative effects around 400 ms, followed by greater P600 effects. Two types of negative effects around 400 ms have been reported in response to a range of morphosyntactic violations (Molinaro, Barber, & Carreiras, 2011; see Caffarra, Mendoza, Davidson, 2019 for a discussion on how distinguishable these two effects are): N400

---

[1] Besides allocutive agreement, Basque has no grammatical gender system.

(centro-posteriorly distributed) and LAN (left-anteriorly distributed) effects. Previous ERP studies on agreement violations similar to those examined here (person agreement violations in Basque) have consistently shown N400 effects followed by P600 effects (Mancini, Massol, Duñabeitia, Carreiras, & Molinaro, in press; Zawiszewski, Santesteban, & Laka, 2016; see also Mancini, Molinaro, Rizzi, & Carreiras, 2011a for similar effects elicited by person anomalies in Spanish, and Molinaro et al. 2011 for a review), indicating an early stage of analysis possibly relying on information at the interface between morphosyntax and semantics/discourse followed by late controlled processes where higher-level information is considered (Molinaro et al., 2011).

One- and two-step proposals (from the parsing and the referential communication literature) have different predictions regarding allocutive violations. One-step models predict that addressee congruence effects should arise at an early stage of processing, without any principled delay as compared to the time course of morphosyntactic analysis (Clark & Carlson, 1981; Jackendoff, 2002; Hanna et al., 2003; Marsel- Wilson & Tyler, 1980; McDonald et al., 1994; Tanenhaus & Trueswell, 1995). This early integration should appear within the first 500 ms after stimulus onset, as seen in previous ERP and eye-tracking studies (Bornkessel-Schlesewsky et al., 2013; Brown-Schmidt et al., 2008; Hanna et al., 2003; Hanna & Tanenhaus, 2004; Hanulíková & Carreiras, 2015; Heller et al., 2008; Nadig & Sedivy, 2002; Van Berkum et al., 2008). In contrast, two-step models (Fodor, 1983; Frazier, 1987; Frazier & Clifton, 1996; Friederici, 2002; Forster & Ryder, 1971; Horton & Keysar, 1996; Keysar et al., 1998) predict that addressee congruence effects should be visible only at a later stage of processing (Keysar et al., 2000; Keysar et al., 2003; Lattner and Friederici, 2003), presumably resulting in a P600 effect (similar to Lattner and Friederici, 2003).

Finally, as an exploratory comparison it is interesting to look at ERP differences between masculine and feminine allocutive forms. Previous behavioral findings showed a difference in the way allocutive forms are processed, with more disruptive effects for the incorrect use of feminine as compared to masculine forms (Wolpert, Mancini, & Caffarra, 2017), which might be related to frequency differences (masculine forms of allocutive are used more often than feminine forms, including sometimes for female addressees; Echeverria, 2003). These disruptive effects will be possibly

8

reflected by the ERP waveforms with a consequent effect of allocutive gender. However, it is not possible to anticipate the time window and the direction of this effect as no ERP study has tested allocutive gender differences before.

2 Method

2.1 Participants

Twenty-five Basque native speakers participated in the experiment (10 women; mean age: 30 y; SD: 7.3). All participants were proficient with Basque allocutive agreement and used it on a weekly basis. They reported normal or corrected-to normal vision and they had no auditory or neurological disorders. Table 1 summarizes the scores from a language background questionnaire and a proficiency test on Basque allocutive agreement. Participants reported high levels of proficiency in Basque *batua* (standard version of the Basque language, commonly used in education and in written Basque). They were exposed to allocutive forms early in life, and consistently used allocutive agreement with close friends or relatives. On the allocutive proficiency test, they showed intermediate-to-high levels of performance, with no difference between masculine and feminine forms ($t(20)=1.45$, $p=.16$). However, they reported producing and hearing masculine forms of allocutive more often. In addition, they reported conversations between two men to be the most typical situation for allocutive usage.

---Table1---

Note that, as all participants were from the Spanish side of the Basque Country, there were all Basque-Spanish early bilinguals (AoAs<10 years old for both languages).

2.2 Materials

Auditory conversations were adapted from Wolpert et al. (2017) using Basque *batua*. All stimuli were recorded at 44.1 kHz and normalized to the average root mean squared amplitude. The audio editing was carried out using GoldWave v6.13. All speakers were Basque native speakers who were highly-proficient allocutive users (n=3, 1 female; mean age: 27.7, SD: 0.6; Basque AoA: 0). Basque conversations included two interlocutors (named Person A and Person B) using allocutive forms.

9

Person A spoke first and could be male or female. Person B replied to Person A and was always male. Person B's reply contained an auxiliary verb (target verb) that either agreed or disagreed with Person A's gender. The target verb could be either masculine or feminine allocutive. This resulted in four experimental conditions (summarized in Table 2): allocutive masculine correct, allocutive masculine incorrect, allocutive feminine correct, allocutive feminine incorrect (38 trials per condition). Person A's and B's utterances were recorded separately. Correct and incorrect conditions were subsequently obtained by cross-splicing each version of Person A's utterance (from a male or a female voice) with Person B's target utterance. In this way, the same target utterance belonged to either the correct or the incorrect condition depending on the previous context. Feminine and masculine versions of the target verbs were matched in duration (masculine: 459 ms, SD: 142; feminine: 438 ms, SD: 140; $t(302)=1.29$, $p>.05$) and word frequency (masculine: 60, SD: 120; feminine: 65, SD: 223; $t(302)<1$, $p>.05$; extracted from the Euskal Hiztegiaren Maiztasun Egitura database; Acha, Laka, Landa & Salaburu, 2014[2]).

As a control comparison, additional auditory conversations were included with a similar structure to those described above. In this subset, Person B's utterances could be grammatically correct or contain a subject-verb person agreement violation (30 trials per condition, see Table 2). Half of Person A's utterances were produced by a male voice, and half were produced by a female voice. Agreeing and disagreeing verbs were matched in duration (correct: 441 ms, SD: 152; incorrect: 481 ms, SD: 136; $t(118)=1.50$, $p>.05$) and word frequency (correct: 29826, SD: 71275; incorrect: 1313, SD: 3137; $t(62)=1.93$, $p>.05$).

Note that in the person manipulation the target utterances' grammaticality was defined regardless of interlocutor gender. This was not true for the allocutive manipulation. In both cases, there was at least one filler word between the target verbs and the end of the utterances (average duration: 838 ms, SD: 280) to avoid wrap-up effects. The overall durations of target verbs used in allocutive and person

---

[2] A Basque speech corpus (Goenkale Corpusa, https://www.ehu.eus/es/web/eins/goenkale-corpusa) was also used to extract the frequency of masculine and feminine verbs. Although not all experimental verbs were present in this Corpus, an analysis on the remaining verbs did not show frequency differences.

10

manipulations did not differ (person: 454 ms, SD: 142; allocutive: 461 ms, SD: 145; $t(210)<1$, $p=.43$)

and in both cases the mismatching information appeared at the end of the target verb. Hence, the time

at which person and allocutive violations became available did not differ. Person's B reply could be

affirmative or negative to make the materials more ecological. Affirmative and negative responses

were counterbalanced across correct and incorrect conditions, and the proportions of affirmative and

negative utterances were similar across violation type (Person: 11.7%; Allocutive: 12.5%; $\chi 2<1$;

$p=.98$).

---Table 2---

2.3 Procedure

Participants were seated in a dark and silent room. They were instructed to listen carefully to the

Basque conversations and rate their naturalness from one (not natural at all) to seven (very natural).

This task has been reported to be more sensitive to the allocutive manipulation as compared to a

grammaticality judgment task (Wolpert et al., 2017).

Each trial began with the presentation of a fixation cross in the center of the screen, and participants

were encouraged to blink during this time. After 1200 ms, the conversations were auditorily presented

while the fixation cross remained on the screen (average duration of the auditory presentation: 8 sec,

SD: 1). After the conversation offset, a 300-ms blank was presented and a prompt for a naturalness

rating appeared on the screen with a scale from one (not natural at all) to seven (very natural). The

question remained on the screen until participants' response or for a maximum of seven seconds.

The experimental session was divided in two blocks (of 106 trials each; about 15 minutes) and was

preceded by a practice session (six trials) to familiarize the participants with the task. The

experimental trials included intermixed person and allocutive manipulations (no fillers were

included). This variability, together with the use of online behavioral measures, enabled us to identify

and exclude any participant using superficial strategies (e.g., focusing only on one type of violation).

At the end of the recording session, the participants performed a proficiency test on Basque allocutive

and filled out a language background questionnaire.

2.4 EEG recording

The electroencephalogram (EEG) was recorded from 27 Ag/AgCl electrodes placed in an elastic cap:

Fp1, Fp2, F7, F8, F3, F4, FC5, FC6, FC1, FC2, T7, T8, C3, C4, CP5, CP6, CP1, CP2, P3, P4, P7, P8,

O1, O2, Fz, Cz, Pz. Additional external electrodes were placed on the mastoids (A1, A2) and around

the eyes (Ve1, Ve2, He1, He2). All sites were online referenced to the left mastoid (A1). Data were

recorded at a sampling rate of 500 Hz. Impedance was kept below 5 KΩ for scalp electrodes, and 10

KΩ for the external electrodes. EEG recordings were re-referenced off-line to the average activity of

the left and right mastoid. The data were filtered offline with a bandpass of 0.01-30 Hz. Horizontal

and vertical eye movements were corrected following the ICA procedure. We decomposed the whole

electroencephalogram in Independent Components for each subject. Then we focused on the

components that explained the highest percentage of the variance in the vertical and horizontal

oculograms. The time course of these components was visually inspected (to ensure they represented

real artifacts) and subtracted from the original data. For each target verb, an epoch of 1700 ms was

obtained including a 200-ms prestimulus baseline. Each epoch was time locked to the onset of the

target verb and was included in further analysis only if associated with a correct behavioral response

at the naturalness judgement (for each trial responses from one to three were correct for incorrect

conditions and responses from five to seven were correct for control conditions). Muscular artifacts

exceeding ±70 μV in amplitude were rejected. Data from four participants were excluded for each

comparison of interest (person and allocutive) since the percentage of artifact-free correct trials was

less than 50% of the overall number of trials. For the remaining participants (n=21, 10 women), an

average of 72% of trials were considered for further analyses. There were no differences in the

number of remaining trials across conditions (masculine correct: 82%, SD: 12, masculine incorrect:

75%, SD: 19; feminine correct: 79%, SD: 12, feminine incorrect: 81%, SD: 16; Allocutive:

$F(3,80)=1.01$, $p=.39$; person correct: 73%, SD: 20, person incorrect: 82%, SD: 17; Person:

$F(1,40)=2.75$, $p=.11$). Average ERP waveforms were computed for each condition.

12

2.5 Statistical analyses

All statistical analyses were carried out using the R statistical package (R Core Team, 2018). For the behavioral data, naturalness ratings were considered only for correct responses. The allocutive results were analyzed by subject and by item with a 2x2 repeated measures analysis of variance (ANOVA) with Agreement (Correct and Incorrect) and Allocutive Gender (Masculine and Feminine) as within-subject and within-item factors. The person results were analyzed by subject and by item with a one-way repeated measures ANOVA with Agreement (Correct and Incorrect) as a within-subject and within-item factor.

For the EEG data, statistical analyses were carried out on three successive time windows selected based on 50 ms-by-50 ms analysis (250-400 ms; 500-800 ms; 800-1100 ms; see Figure 1 for further details on the time windows selection). ERP effects were evaluated taking into account nine groups of electrodes. Each group represented the mean amplitude of three electrodes in proximate positions: Left Anterior (F3, F7, FC5), Left Central (C3, T7, CP5), Left Posterior (P3, P7, O1), Medial Anterior (Fp1, Fp2, Fz), Medial Central (FC1, FC2, Cz), Medial Posterior (CP1, CP2, Pz), Right Anterior (F4, F8, FC6), Right Central (C4, T8, CP6), Right Posterior (P4, P8, O2). These groups were included in the statistical analyses as different levels of two topographical factors: Longitude (Anterior, Central and Posterior) and Latitude (Right, Medial and Left). For the allocutive agreement manipulation, a repeated measures ANOVA was performed for each time window of interest with Agreement (Correct, Incorrect), Allocutive (Masculine, Feminine), Longitude (Anterior, Central, Posterior), and Latitude (Left, Medial, Right) as within-subject factors. For the person agreement manipulation, a repeated measures ANOVA was performed for each time window of interest including Agreement (Correct, Incorrect), Longitude (Anterior, Central, Posterior), and Latitude (Left, Medial, Right) as within-subject factors. A final ANOVA for each time window of interest was conducted including violation Type (Masculine Allocutive, Feminine Allocutive, Person) and Agreement (Correct, Incorrect) to directly compare the timing of different violation effects. The Greenhouse-Geisser procedure was applied whenever the sphericity assumption was violated and FDR post-hoc corrections were applied to all ANOVAs carried out (Hochberg, 1988). Effects of topographical factors are reported only when they interacted with the experimental factors.

13

---Figure 1---

3 Results

3.1 Behavioral data

Figure 2 shows the naturalness rating results for each condition. For the person manipulation, there was a significant effect of Agreement $(F_1(1,20)=402, p<0.001, \eta_p^2=0.914; F_2(1,30)=197, p<0.001, \eta_p^2=0.78)$, with incorrect items rated an average of 5.5 points less natural than correct items (correct: 6.61, SD: 0.67; incorrect: 1.11, SD: 0.40). For the allocutive manipulation (masculine correct: 6.75, SD: 0.53; masculine incorrect: 1.48, SD: 0.61; feminine correct: 6.75, SD: 0.55; feminine incorrect: 1.26, SD: 0.50), there was a significant effect of Agreement $(F_1(1,20)=914.8, p<0.001, \eta_p^2= 0.92; F_2(1,37)=2349.2, p<0.001, \eta_p^2=0.94)$, with incorrect items rated an average of 5.4 points less natural than correct items (correct: 6.75, SD: 0.54, incorrect: 1.37, SD: 0.57). There was no significant effect of Allocutive Gender $(F_1(1,20)=0.5, p=0.5, \eta_p^2=0.007; F_2(1,37)=1.2, p=0.3, \eta_p^2=0.008)$ and no significant interaction between Allocutive Gender and Agreement $(F_1(1,20)=0.1, p=0.7, \eta_p^2=0.001; F_2(1,37)=0.5, p=0.5, \eta_p^2=0.004)$.

---Figure 2---

3.2 ERPs

Figure 3 shows the ERP average waveforms for each experimental condition. Person violations seem to elicit a greater negativity around 400 ms followed by greater positivity as compared to correct sentences. Allocutive violations only show greater positivities than correct sentences.

--- Figure 3---

3.2.1 Person. No significant effect was reported before 250 ms (all ps>.05).

3.2.1.1 250-400 ms. A main effect of Agreement was found $(F(1, 20)=6.99, p<.05, \eta_p^2=0.26)$, with the incorrect condition being more negative than the correct condition (correct: 0.30 µV, SD: 2.07, incorrect: -0.68 µV, SD: 2.03). No other effect was reported (all ps>.05).

3.2.1.2 500-800 ms. The effect of Agreement was significant ($F(1, 20)$=21.07, p<.001, η_p^2= 0.51) and it also interacted with Latitude ($F(2, 40)$=9.28, p<.001, η_p^2=0.32). The incorrect condition elicited more positive waveforms than the correct condition (correct: 1.75 µV, SD: 2.49, incorrect: 4.32 µV, SD: 3.91), and this effect was stronger over left and medial sites (left: $t(20)$=4.76, p<.001; medial: $t(20)$=4.80, p<.001; right: $t(20)$=3.55, p<.01).

3.2.1.3 800-1100 ms. There was a main effect of Agreement ($F(1, 20)$=13.68, p<.01, η_p^2= 0.41), and significant interactions with the topographical factors (Agreement x Latitude: $F(2, 40)$=4.63, p<.05, η_p^2=0.19; Agreement x Longitude: $F(2, 40)$=15.59, p<.001, η_p^2=0.44). Greater positives were observed for the incorrect condition relative to the correct condition (correct: 1.92 µV, SD: 2.39, incorrect: 3.88 µV, SD: 3.58). This effect was centro-posteriorly distributed (anterior: $t(20)$=1.47, p=.16; central: $t(20)$=3.37, p<.01; posterior: $t(20)$=4.91, p<.001), and greater over left-medial sites (left: $t(20)$=3.65, p<.01; medial: $t(20)$=3.97, p<.01; right: $t(20)$=2.69, p<.05).

3.2.2 Allocutive. ~~No significant effect was reported before 250 ms (all ps>.05).~~

3.2.2.1 250-400 ms. There was no significant effect involving the experimental factors (Agreement: $F(1, 20)$<1, p=.51, η_p^2=0.02; Allocutive Gender: $F(1, 20)$<1, p=.66, η_p^2=0.01; Agreement x Allocutive Gender: $F(1, 20)$<1, p=.88, η_p^2=0.001; all other interactions: Fs<2; ps>.05; masculine correct: -0.61 µV, SD: 1.83, , masculine incorrect: -0.50 µV, SD: 2.18, feminine correct: -0.72 µV, SD: 1.94, feminine incorrect: -0.54 µV, SD: 1.95).

3.2.2.2. 500-800 ms. There was a significant interaction between Agreement and Longitude ($F(2, 40)$=4.72, p<.05, η_p^2=0.19). The FDR corrected post-hoc t-tests showed more positive waveforms over posterior sites than anterior sites (correct: $t(20)$=6.82, p<.001; incorrect: $t(20)$=7.23, p<.001; correct anterior: 0.90 µV, SD: 2.53, correct posterior: 3.60 µV, SD: 2.30, incorrect anterior: 0.86 µV, SD: 2.98, incorrect posterior: 4.11 µV, SD: 2.58). However, no significant difference between the correct and the incorrect condition was reported (anterior: $t(20)$<1, p=.89; central: $t(20)$<1, p=.89; posterior: $t(20)$=1.73, p=.30). No other effect was significant (Agreement: $F(1, 20)$<1, p=.42, η_p^2=0.03; Allocutive Gender: $F(1, 20)$=1.88, p=.19, η_p^2=0.09; Agreement x Allocutive Gender:

$F(1, 20)<1$, $p=.36$, $\eta_p^2=0.04$; all other interactions: $Fs<2$; $ps>.05$; masculine correct: 1.98 µV, SD: 2.58, , masculine incorrect: 2.41 µV, SD: 3.01, feminine correct: 2.56 µV, SD: 2.68, feminine incorrect: 2.56 µV, SD: 3.09).

3.2.2.3 800-1100 ms. The main effect of Agreement ($F(1, 20)=16.20$, $p<.001$, $\eta_p^2=0.45$) and the interactions between Agreement and the topographical factors were significant (Agreement x Longitude: $F(2, 40)=8.62$, $p<.01$, $\eta_p^2=0.30$; Agreement x Longitude x Latitude: $F(4, 80)=3.99$, $p<.05$, $\eta_p^2=0.17$). Greater positivities were reported for the incorrect condition than the correct condition (correct: 1.96 µV, SD: 2.58, incorrect: 3.23 µV, SD: 3.37), with a centro-posterior distribution (anterior sites: $ps\geq.05$; left central: $t(20)=4.99$, $p<.001$; medial central: $t(20)=3.37$, $p<.01$; right medial: $t(20)=2.71$, $p<.05$; left posterior: $t(20)=4.44$, $p<.001$; medial posterior: $t(20)=5.24$, $p<.001$; right posterior: $t(20)=4.81$, $p<.001$).

The main effect of Allocutive Gender was also significant ($F(1, 20)=7.63$, $p<.05$, $\eta_p^2=0.28$), with greater positivities for the feminine than the masculine allocutive (masculine: 2.31 µV, SD: 3.21, feminine: 2.88 µV, SD: 2.90).

Importantly, the interaction between Agreement and Allocutive Gender was significant ($F(1, 20)=11.45$, $p<.01$, $\eta_p^2=0.36$). Masculine allocutive elicited a stronger violation effect as compared to feminine allocutive (masculine: 1.85 µV, $t(20)=4.72$, $p<.001$; feminine: 0.71 µV, $t(20)=2.18$, $p<.05$). This difference was due to the correct condition (masculine vs. feminine: $t(20)=4.90$, $p<.001$; masculine: 1.39 µV, SD: 2.56, feminine: 2.52 µV, SD: 2.49), while the incorrect conditions did not differ across allocutive gender (masculine vs. feminine: $t(20)<1$, $p=.98$; masculine: 3.24 µV, SD: 3.52, feminine: 3.23 µV, SD: 3.22; see Figure 4).

Significant interactions with Agreement, Allocutive Gender and topographical factors were also found (Agreement x Allocutive Gender x Latitude: $F(2, 40)=3.39$, $p<.05$, $\eta_p^2=0.14$; Agreement x Allocutive Gender x Longitude x Latitude: $F(4, 80)=6.41$, $p<.001$, $\eta_p^2=0.24$). FDR-corrected post-hoc t-tests revealed that the P600 effect was posteriorly distributed for feminine allocutive (posterior and left central sites: $ts>2$, $ps<.05$; other sites: $ts<2$, $ps>.05$) and more broadly distributed for masculine

allocutive (left medial anterior sites: $ts>2$, $ps<.05$; right anterior site: $t<2$, $p>.05$; central sites: $ts>3$,

$ps<.01$; posterior sites: $ts>4$, $ps<.001$). This different distribution was due to the agreeing allocutive

forms. Within the correct condition, feminine allocutive showed more broadly distributed positivities

than masculine allocutive (anterior: $ts>2$, $ps<.05$; central: $ts>3$, $ps<.01$; left and medial posterior: $ts>3$,

$ps<.01$; right posterior: $t=2.07$, $p=.05$)[3]. In the incorrect condition, the positive waveforms for

masculine and feminine allocutive were similarly distributed ($Fs<1$, $ps>.05$; see Figure 3 and 4).

---Figure 4---

To summarize, person agreement violations elicited a greater broadly distributed negativity (250-400

ms) followed by a greater P600 (500-1100 ms) as compared to correct utterances. Allocutive

violations elicited a greater P600 effect in a late time window (800-1100 ms) as compared to the

grammatical counterparts. Greater P600 effects were reported for violations of masculine allocutive as

compared to feminine allocutive. However, the distinct size of the effect was due to differences in the

correct condition.

Finally, a direct comparison among the three agreement violation types (Masculine Allocutive,

Feminine Allocutive, Person) showed that an effect of person violations was present earlier than the

effect of allocutive violations (significant interactions between Type and Agreement in all time

windows of interest: 250-400: $F(2,40)=3.27$, $p<.05$, $\eta_p^2=0.14$; 500-800: $F(2,40)=10.82$, $p<.001$, $\eta_p^2=$

0.35; 800-1100: $F(2,40)=3.60$, $p<.05$, $\eta_p^2=0.15$; see Table 3 for the FDR-corrected follow-up tests

and Figure 5)[4].

---Table 3---

---Figure 5---

---

[3] Additional analyses showed that the sex of the participants did not modulate the allocutive violation effect.
[4] Results from the analysis of ERP responses time-locked to the offset of the target verbs showed a similar shift in time between the two types of violations.

4 Discussion

The present study was aimed at exploring the role of addressee-related information during speech comprehension. Auditory dialogues were constructed so that the presence of grammatical violations could be detected based on an addressee identity incongruence (allocutive violations) or on a morphosyntactic mismatch (person agreement violations). The results revealed that violations depending on pragmatic sources of information (addressee gender) elicited ERP effects starting from a relatively late stage of analysis (800 ms after the onset of the target verb), with a greater P600 for the grammatically incorrect as compared to the correct condition. However, grammatical violations depending on morphosyntactic incongruences elicited earlier ERP effects, with a greater N400[5] followed by a greater P600 response as compared to the correct condition (in line with previous ERP results on person agreement violations: Mancini et al., 2011a; 2011b; 2016; Zawiszewski et al., 2016; for a review see Molinaro et al., 2011).

The present findings represent a case where discourse information does not seem to affect the time course of speech comprehension at early stages of processing, or at least not as early as morphosyntactic information does. These results are compatible with the idea that pragmatic sources of information are integrated at a late stage of analysis (Fodor, 1983; Frazier & Clifton, 1996; Friederici, 2002; Forster & Ryder, 1971; for similar proposals see also: Horton & Keysar, 1996; Keysar et al., 1998; Pickering & Garrod, 2004). Overall, they are in line with theoretical proposals assuming encapsulated modules (where different linguistic components initially work separately), and provide important theoretical constraints for interactive views (where different sources of linguistic information interact at the earliest stages of analysis).

The present findings represent an exception compared to a number of eye-tracking and electrophysiological studies showing that interlocutor-dependent information interacts at an early

---

[5] Note that although the interpretation of the N400 effect is still highly debated (Friederici, 2002; Kutas & Federmeier, 2000), the crucial focus of our argument is that the N400 effect reported here (1) must be related to how the brain is initially treating morphosyntactic information that is not interlocutor-related, and (2) is temporally localized at an earlier stage of analysis as compared to our pragmatic effects.

18

stage of analysis with other sources of non-pragmatic information (Bornkessel-Schlesewsky et al.,

2013; Brown-Schmidt et al., 2008; Hanna et al., 2003; Hanna & Tanenhaus, 2004; Hanulíková &

Carreiras, 2015; Heller et al., 2008; Nadig & Sedivy, 2002; Van Berkum et al., 2008). This pattern of

results is difficult to reconcile with a strict version of one-step models from referential communication

stating that any available interlocutor-related information has a rapid impact on language

comprehension and automatically updates the sentence interpretation based on common ground

knowledge (Clark & Carlson, 1981; Galati & Brennan, 2010; Hanna, Tanenhaus & Trueswell, 2003;

Jackendoff, 2002; Tanenhaus & Trueswell, 1995). Note that the present findings do not necessarily

rule out any one-step account, but they surely provide new insights about constraints that need to be

considered within these theoretical proposals.

Part of the disagreement between the present findings and previous research evidence might be

accounted for by the type of discourse information taken into account in each study and its relevance

relative to the conversation topic and the experimental task. It is worth noting that previous

electrophysiological studies have been focused on the role of the speaker, rather than the addressee. In

these studies, the content of the experimental utterances consisted of speakers' preferences, opinions,

and thoughts, making the speaker perspective pivotal for sentence understanding. On the other hand,

previous eye-tracking studies that instead focused on the role of the addressee employed collaborative

tasks where the knowledge shared between speaker and addressee was essential to successfully

complete the task. Most of the studies adopting these paradigms provided evidence supporting early

availability and integration of interlocutor-based information (but see Lattner & Friederici, 2003;

Keysar et al., 2000; 2003). The high relevance of the discourse information under study might have

contributed to its early impact on sentence comprehension processing.

Unlike previous experimental designs, the present study utilized dialogues where the discourse level

information was less crucial for the correct execution of the task (our naturalness rating provided a

range of possible responses where more than one option could be considered correct). The

experimental manipulation was not focused on the protagonist of the speech act and the addressee-

related information of interest (i.e., gender) was not associated with the main topic of the

conversation. Note that the above considerations do not imply that addressee-based features were ignored within the present experimental design (the naturalness judgement showed that incorrect allocutive forms were clearly detected and judged as unnatural as person agreement violations).

One possible explanation of present and previous findings might be that the time point when discourse information is integrated in sentence comprehension depends on its degree of relevance relative to the conversational settings and/or to the task requirements. This proposal is in line with theoretical models of sentence comprehension assuming that the number of processing stages (one or two steps) can change as a function of a number of constraints, such as information salience and relevance (Marslen-Wilson & Tyler, 1980). According to these models, the integration of discourse information can flexibly start at different points in time (early or late stages of processing) and be speeded up based on probabilistic cues (Marslen-Wilson & Tyler, 1980; McDonald et al., 1994). This would suggest that, under specific circumstances, a parallel interactive system can be reduced to a serial one (Marslen-Wilson & Tyler, 1980). Hence, one- and two- step models would be two possible outcomes of the same continuum. Within this framework, the present results would suggest that the low semantic association between addressee-based information and utterance meaning (see Tabossi, 1988 for a similar proposal) might delay the integration of discourse level information and stretch the distance between the non-pragmatic and the pragmatic stages of analysis. This exciting possibility needs further testing to be confirmed.

Another interesting aspect of the present results regards the ERP differences between masculine and feminine allocutive. Although the brain waveforms to the violation of both types of allocutive elicited a similar P600 effect, the voltage responses to the grammatical condition differed. Grammatically correct feminine allocutives elicited greater broadly-distributed positivity as compared to grammatically correct masculine allocutives. This was an unexpected result and could not be anticipated based on the previous literature. No ERP study had tested this comparison before, and behavioral findings showed an effect of allocutive gender in the violation condition rather than in the control one (Wolpert et al., 2017). As a result, this effect must be treated as an exploratory finding and interpreted with caution. Here below we consider a few post-hoc tentative explanations.

20

The ERP difference observed in the agreement condition might be related to different patterns of usage/frequency of exposure between the two types of allocutive. On one hand, participants seemed to have a different frequency of exposure to feminine and masculine allocutive in their daily life. For instance, our participants reported masculine allocutive to be more used as compared to feminine allocutive (in line with previous literature, Echeverria, 2003; Wolpert et al., 2017). On the other hand, there was a difference in frequency of exposure within the experimental section as well, with masculine agreeing allocutive forms being heard more often than feminine agreeing allocutive forms (as Person B was always male). Previous ERP studies have reported amplitude modulation of P600 responses as a function of the frequency of exposure to a specific grammatical construction, with greater P600 responses for less frequent constructions (see Hahne & Friederici, 1999 where the frequency of exposure changed within the experimental session; see Caffarra & Martin, 2019, where frequency of exposure differences were present on a daily basis). Similarly, the reported ERP differences between female and masculine allocutive might reflect distributional differences (within participants' daily life and/or within the experimental session) which were specific to the agreement condition.

Another possible explanation is that the P600 difference observed between congruent allocutive forms would reflect (at least partially) distinct cognitive mechanisms as compared to the P600 difference observed for allocutive violations. This seems to be supported by the presence of a different topographic distribution for the two ERP effects (centro-parietal for allocutive violation; broadly distributed for agreeing allocutive gender type), possibly indicating the involvement of different neural sources. Previous ERP studies on parsing have pointed out that while posterior-distributed P600 effects are associated with reanalysis and repair processes, more anterior topographic distribution can be related to processing costs due to increasing complexity at the discourse level (Kaan & Swaab, 2003). The use of allocutive in a man-to-woman dialogue did represent a more pragmatically complex situation, as this choice can convey additional information about the speaker's intended meaning (e.g., feminine allocutive is in danger of being lost and its use with female addresses, even if they are not closely related to the speaker, might signal the wish of the speaker to

21

maintain its use; Hualde, Lakarra, Trask, 1995). However, this represents a tentative explanation which needs further testing on the role of speaker identity in within (male-male, female-female) and between (male-female, female-male) gender-based groups.

Finally, the present findings confirmed that P600 effects can be reported with a wide range of linguistic manipulations, from grammatical violations based on morphosyntactic mismatches (e.g., person agreement violations; Molinaro et al, 2011) to those that depend on pragmatic sources of information (e.g., allocutive violations; see also Regel, Coulson, & Gunter, 2010; Regel, Gunter, Friederici, 2011). Interestingly, while the first type of violations elicited a longer lasting P600 effect, which can be possibly divided into two phases (early, broadly distributed positivity followed by a more centro-posterior effect; as already reported in previous ERP studies; see Molinaro et al., 2011 for a review), the second type of violation elicited only a late centro-posterior positivity. This dissociation seems to fit well with the theoretical proposal that the first phase of the P600 would reflect integration difficulties between the target constituent and the preceding sentence fragment, while the later phase would be more associated with reanalysis processes involving higher-level information (Barber & Carreiras, 2005; Friederici, Mecklinger, Spencer, Steinhauer, Donchin, 2001; Hagoort & Brown, 2000). However, the present findings are not compatible with the idea that discourse-level information specifically affects the first phase of analysis (cfr. Molinaro et al., 2011). This is also partially in line with recent ERP findings showing a difference between syntax-related and pragmatics-related P600 effects (Regel, Meyer & Gunter, 2014), although the modulation that we report here regards not only the topographic distribution but also the timing of the effect.

As a final caveat, it should be noted that all participants involved in this study were early bilinguals, who might show different ERP correlates from monolinguals and other bilingual profiles (Caffarra, Molinaro, Davidson, & Carreiras, 2015, for a review). Hence, additional studies are needed in order to test whether the present results can be generalized to other linguistic profiles.

22

In conclusion, the present ERP study represents a case where interlocutor-based information impacts sentence processing only at a late stage of analysis, when morphosyntactic processing has been already initiated. These results point to theoretical constraints for one-step accounts.

23

References

Acha, J., Laka, I., Landa, J. & Salaburu, P. (2014). EHME: A new word database for research in

Basque language. *The Spanish Journal of Psychology*, 17, E79. doi:

https://www.doi.org/10.1017/sjp.2014.79

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP

comparison. *Journal of Cognitive Neuroscience*, 17(1): 137-153. doi:

https://www.doi.org/10.1162/0898929052880101

Bornkessel-Schlesewsky, I., Krauspenhaar, S., & Schlesewsky, M. (2013). Yes, you can? A speaker's

potency to act upon his words orchestrates early neural responses to message-level meaning.

PLoS One, 24, e69173. doi: https://www.doi.org/10.1371/journal.pone.0069173

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010).Two minds, one dialog: coordinating speaking and

understanding. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 301-

344). Burlington: Academic Press.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from

private information when interpreting questions during interactive conversation. *Cognition*,

107, 1122-1134. doi: https://www.doi.org/10.1016/j.cognition.2007.11.005

Bühler, K. L. (1982). The deictic field of language and deictic words. In R. J. Jarvella and Wolfgang,

K. (Eds.), *Speech, Place and Action: Studies in deixis and related topics* (pp. 9-30).

Chichester: John Wiley & Sons, Ltd.

Caffarra, S., & Martin, C. (2019). Not all errors are the same: ERP sensitivity to error typicality in

foreign accented speech perception. *Cortex*, 116, 308-320. doi:

https://www.doi.org/10.1016/j.cortex.2018.03.007

Caffarra, S., Mendoza, M., & Davidson, D. (2019). Is the LAN effect in morphosyntactic processing

an ERP artifact? *Brain & Language*, 191, 9-16. doi:

https://www.doi.org/10.1016/j.bandl.2019.01.003.

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic

processing revealed through event-related potentials: An empirical review. *Neuroscience &*

Biobehavioral reviews, 51C, 31-47. doi:

https://www.doi.org/10.1016/j.neubiorev.2015.01.010.

Clark, H. H. & Carlson, T. B. (1981). Context form comprehension. In J. Long and A. Baddeley

(Eds.), *Attention and performance*, 9, (pp. 313-330). Hillsdale, NJ: Erlbaum.

Echeverria, B. (2003). Language ideologies and practices in (en)gendering the Basque nation.

Language in Society, 32(03), 383-413. doi:

https://www.doi.org/10.1017/S0047404503323048

Echeverria, B. (2010). For whom does language death toll? Cautionary notes from the Basque case.

Linguist. Edu. 21, 197–209. doi: https://www.doi.org/10.1016/j.linged.2009.10.001

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, K. I. & Ryder, L. A. (1971). Perceiving the structure and meaning of sentences. *Journal of

Verbal Learning and Verbal Behavior*, 10, 285-296. doi: https://www.doi.org/10.1016/S0022-

5371(71)80056-7

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and

performance: vol.12. The Psychology of Reading* (pp. 559-586). Hove, UK: Erlbaum.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: The MIT Press.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive

Sciences*, 6, 78-84. doi: https://www.doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic

parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related

brain potentials. *Cognitive Brain Research*, 11, 305-323. doi:

https://www.doi.org/10.1016/S0926-6410(00)00065-3

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the

speaker, or for the addressee? *Journal of Memory and Language*, 62, 35-51. doi:

https://www.doi.org/10.1016/j.jml.2009.09.002

Glendinning, L. (2008, November 6). Obama is young, handsome and tanned, says Silvio Berlusconi.

The Guardian. Retrieved from http://www. theguardian.com

25

Haddican, W. F. (2003). Dialect contact in a Southern Basque town. Lang. Var. Change 15, 1–35. doi:

https://www.doi.org/10.1017/S0954394503151010

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic

analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*,

11(2), 194-205.

Hagoort, P., & Brown, C., M. (2000). ERP effects of listening to speech compared to reading: the

P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation.

Neuropsychologia ,38, 1531-1549. doi: https://www.doi.org/10.1016/S0028-3932(00)00053-1

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative

task: evidence from eye movements. *Cognitive Science*, 28, 105-115. doi:

https://www.doi.org/10.1207/s15516709cog2801_5

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and

perspective on domains of referential interpretation. *Journal of Memory and Language*, 49,

43-61. doi: https://www.doi.org/doi.org/10.1016/S0749-596X(03)00022-6

Hanulíková, A., & Carreiras, M. (2015). Electrophysiology of subject-verb agreement mediated by

speakers' gender. *Frontiers in Psychology*, 6, 1396.doi:

https://www.doi.org/10.3389/fpsyg.2015.01396

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains

of reference. *Cognition*, 108, 831-836. doi:

https://www.doi.org/10.1016/j.cognition.2008.04.008

Hochberg, B. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*,

75(4), 800-802. doi: https://www.doi.org/10.1093/biomet/75.4.800

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*,

59, 91-117.

Hualde, J. I., Lakarra, J., A., Trask, R. L. (1995). Towards a history of Basque language. Amsterdam/

Philadelphia: John Benjamins.

Hualde, J. I., & Ortiz de Urbina, J. (2003). *A grammar of Basque*. Berlin: Mouton de Gruyter.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford, UK:

Oxford University Press.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: an

electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15, 98-110. doi:

https://www.doi.org/10.1162/089892903321107855

Keysar, B., Barr, D. J. & Horton, W. S. (1998). The egocentric basis of language use: Insights from a

processing approach. *Current Directions in Psychological Science*, 7(2), 46-49.

doi:10.1111/1467-8721.ep13175613

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The

role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32-37. doi:

https://www.doi.org/10.1111/1467-9280.00211

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25-41.

Kutas, M., Federmeier K.D. (2000). Electrophysiology reveals semantic memory use in language

comprehension. *Trends in Cognitive Sciences*, 4, 463-470. doi:

https://www.doi.org/10.1016/S1364-6613(00)01560-6

Lattner, S. & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory

sentence processing--evidence from event-related brain potentials. *Neuroscience Letters*, 339,

191-194. doi: https://www.doi.org/10.1016/S0304-3940(03)00027-2

MacDonald, M. C., Pearlmutter, N. J., Seidenberg, M. S. (1994). The lexical nature of syntactic

ambiguity resolution. *Psychological Review*, 101, 676-703. doi:

https://www.doi.org/10.1037/0033-295X.101.4.676

Mancini, S., Massol, S., Duñabeitia, J. A., Carreiras, M., & Molinaro, N. (in press). Agreement and

illusion of disagreement: An ERP study on Basque. *Cortex*.

Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011a). A person is not a number: discourse

involvement in subject-verb agreement computation. *Brain Research*, 1410, 64-76.

doi:10.1016/j.brainres.2011.06.055

27

Mancini, S., Molinaro, N., Rizzi, L. & Carreiras, M. (2011b). When persons disagree: an ERP study of Unagreement in Spanish. *Psychophysiology*, 48(10), 1361-1371. doi: https://www.doi.org/10.1111/j.1469-8986.2011.01212.x.

Marslen- Wilson, W. D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71. doi: https://www.doi.org/10.1016/0010-0277(80)90015-3

Molinaro, N., Barber, H., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 30, 1-2 3. doi: https://www.doi.org/10.1016/j.cortex.2011.02.019

Nadig, A. S., Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*,13, 329-36. doi: https://www.doi.org/10.1111/j.0956-7976.2002.00460.x

Oyharçabal, B. (1993). "Verb agreement with non-arguments: on allocutive agreement," in Generative Studies in Basque Linguistics, eds J. I. Hualde and J. O. de Urbina (Amsterdam: John Benjamins), 89–114.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169-190. doi: https://www.doi.org/10.1017/S0140525X04000056

R Core Team (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Regel, S., Coulson, S., Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, 1311, 121-135. doi: https://www.doi.org/10.1016/j.brainres.2009.10.077

Regel, S., Gunter, T.C., Friederici, A.D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, 23, 277-293. doi: https://www.doi.org/10.1162/jocn.2010.21411

Regel, S., Meyer, L., Gunter, T. (2014). Distinguishing processes reflected by P600 effects: evidence from ERPs and neural oscillations. *PLoS One*, 9, e96840. doi: https://www.doi.org/10.1371/journal.pone.0096840

28

Steinhauer, K. & Drury, J. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120, 135-162. doi: https://www.doi.org/10.1016/j.bandl.2011.07.001

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, 27, 324-340. doi: https://www.doi.org/10.1016/0749-596X(88)90058-7

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Vol. 11. Speech, language, and communication* (pp. 217-262). San Diego, CA: Academic Press.

Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580-91. doi: https://www.doi.org/10.1162/jocn.2008.20054

Wolpert, M., Mancini, S., & Caffarra, S. (2017). Addressee Identity and Morphosyntactic Processing in Basque Allocutive Agreement. *Frontiers in Psychology*, 8, 1439. doi: https://www.doi.org/10.3389/fpsyg.2017.01439.

Zawiszewski, A., Santesteban, M., & Laka, I. (2016). Phi-features reloaded: An event-related potential study on person and number agreement processing. *Applied Psycholinguistics*, 37, 601-626.   doi: https://www.doi.org/https://www.doi.org/10.1017/S014271641500017X

29

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Author Notes

30

Offline Questionnaire	mean	SD	Range
Basque age of acquisition (AoA)	0		
Basque *batua* formal instruction (years)	13.4	6.5	4-23
Basque *batua* production proficiency (%)	86.9	20.3	60-100
Baque *batua* comprehension proficiency (%)	98.6	3.2	90-100
AoA of Basque allocutive (years)	2.1	2.8	0-6
Allocutive proficiency (production, %)	77.6	19.7	50-100
Allocutive proficiency (comprehension, %)	95.5	7.27	70-100
Allocutive usage on weekly basis (%)	33.3	26.7	5-80
Masculine allocutive usage on weekly basis (%)	61.5	33.4	10-100
Feminine allocutive usage on weekly basis (%)	29.3	27.9	1-100
Participants more exposed to masculine allocutive (%)	85.7		
Participants mainly using allocutive to talk to men (%)	81.0		
Participants reporting masculine allocutive as the typical form (%)	71.4		

Allocutive proficiency test	mean	SD	Range
Overall score (%)	76.3	14.7	63.3-100
Masculine allocutive (%)	78.4	16.1	61.5-100
Feminine allocutive (%)	74.3	16.1	50-100

Table 1. Average and SD scores from the offline questionnaire (self-reported measures) and a

Basque proficiency test (sentence completion task where participants had to write the correct

allocutive inflection of 28 verbs).

	Allocutive	
	Correct	**Incorrect**
Masculine	A (male): Etxe berriek bost solairu izango ditiztek ziurrenik. House new$_{pl}$ five floors be$_{FUT}$ AUX $_M$ probably B: Bai, eta nahi duenarentzat, igogailuak izango <u>ditiztek</u>$_M$ garajeetaraino. Yes, and want AUX-nom-for elevators be$_{FUT}$ AUX $_M$ garages-to .	A (female): Etxe berriek bost solairu izango ditiztek ziurrenik. House new$_{PL}$ five floors be$_{FUT}$ AUX $_F$ probably B: *Bai, eta nahi duenarentzat, igogailuak izango <u>ditiztek</u>$_M$ garajeetaraino. Yes, and want AUX-nom-for elevators be$_{FUT}$ AUX$_M$ garages-to
Feminine	A (female): Etxe berriek bost solairu izango ditiztek ziurrenik. House new$_{pl}$ five floors be$_{FUT}$ AUX $_M$ probably B: Bai, eta nahi duenarentzat, igogailuak izango <u>ditizten</u>$_F$ garajeetaraino. Yes, and want AUX-nom-for elevators be$_{FUT}$ AUX $_F$ garages-to.	A (male): Etxe berriek bost solairu izango ditiztek ziurrenik. House new$_{pl}$ five floors be$_{FUT}$ AUX $_M$ probably B: *Bai, eta nahi duenarentzat, igogailuak izango <u>ditizten</u>$_F$ garajeetaraino. Yes, and want AUX-nom-for elevators be$_{FUT}$ AUX $_F$ garages-to.

	Person	
	Correct	**Incorrect**
	A (any): Gaur arratsaldeko gidoigintza ikastaroa izugarria izan da. Today afternoon-of writing course great be AUX B: Bai, halako etorriarekin ikasle berriak hogei bat istorio on idatz <u>ditzake</u> erraz. Yes, such flow-with student new twenty some stories good write AUX$_{3.SG}$ easily	A (any): Gaur arratsaldeko gidoigintza ikastaroa izugarria izan da. Today afternoon-of writing course great be AUX B: *Bai, halako etorriarekin ikasle berriak hogei bat istorio on idatz <u>ditzaket</u> erraz. Yes, such flow-with student new twenty some stories good write AUX$_{1.SG}$ easily

Table 2. English translations, Allocutive correct: A-"The new houses will probably have five floors.", B-" Yes, and for anyone who wishes, there <u>will be</u> lifts to the garages."; Person correct: A-"Today afternoon's writing course was really great.", B-"Yes, with such a brainstorm, the new student <u>will be</u> able to write about twenty good stories.". Person A's gender is defined in

parenthesis. Incorrect sentences are marked with an asterisk. [AUX=auxiliary verb,

NOM=nominalization; FUT=future, 3=3rd person; SG= singular; M=masculine; F=feminine]

	250-400	500-800	800-1100
Feminine Allocutive	$t(20)<1, p=.76$	$t(20)<1, p=.99$	$t(20)=2.18, p<.05$
Masculine Allocutive	$t(20)<1, p=.76$	$t(20)=1.46, p=.32$	$t(20)=4.72, p<.001$
Person	$t(20)=2.64, p<.05$	$t(20)=4.60, p<.001$	$t(20)=3.70, p<.01$

Table 3. FDR-corrected t-tests of the Type x Agreement interaction. Gray cells mark significant results (light gray: <.05; mid gray: <.01; dark gray: <.001). The person violation effect appeared earlier than allocutive effects.
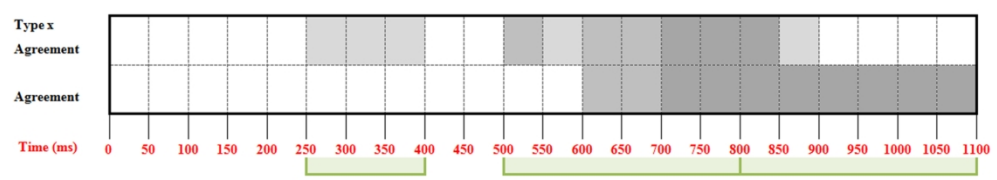
Figure 1. Results from 50 ms-by-50 ms Anovas including Violation Type (Masculine Allocutive, Feminine Allocutive, Person) and Agreement (Correct, Incorrect). Gray colors indicate different p-values thresholds after applying the FDR correction (light: $p<.05$; mid: $p<.01$; dark: $p<.001$). The resulting time windows of interest (including either an effect of Agreement or its interaction with Type) are marked in green. Given the length of the second time window and in order to keep track of the time course of late processes, this segment was divided into two consecutive time windows (500-800; 800-1100 ms).
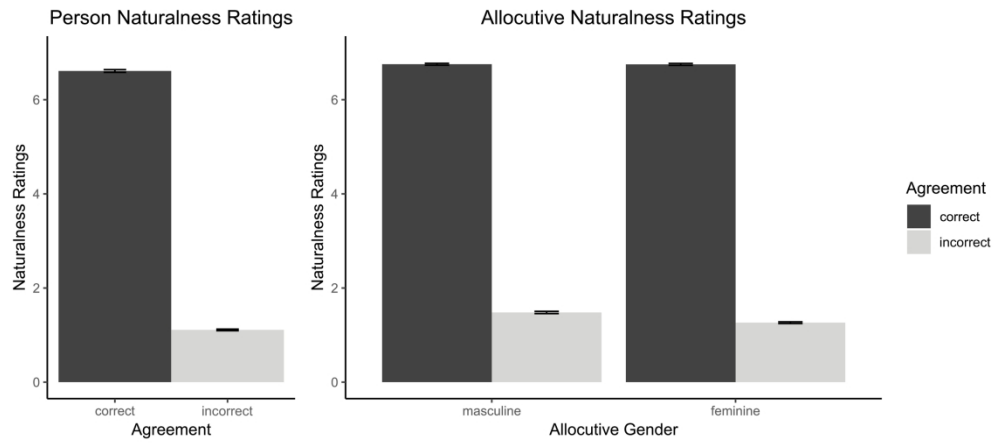
189x36mm (300 x 300 DPI)

Figure 2. Left: Naturalness ratings for items with the person agreement manipulation. Right: Naturalness ratings for items with the allocutive agreement manipulation. Error bars show standard error.
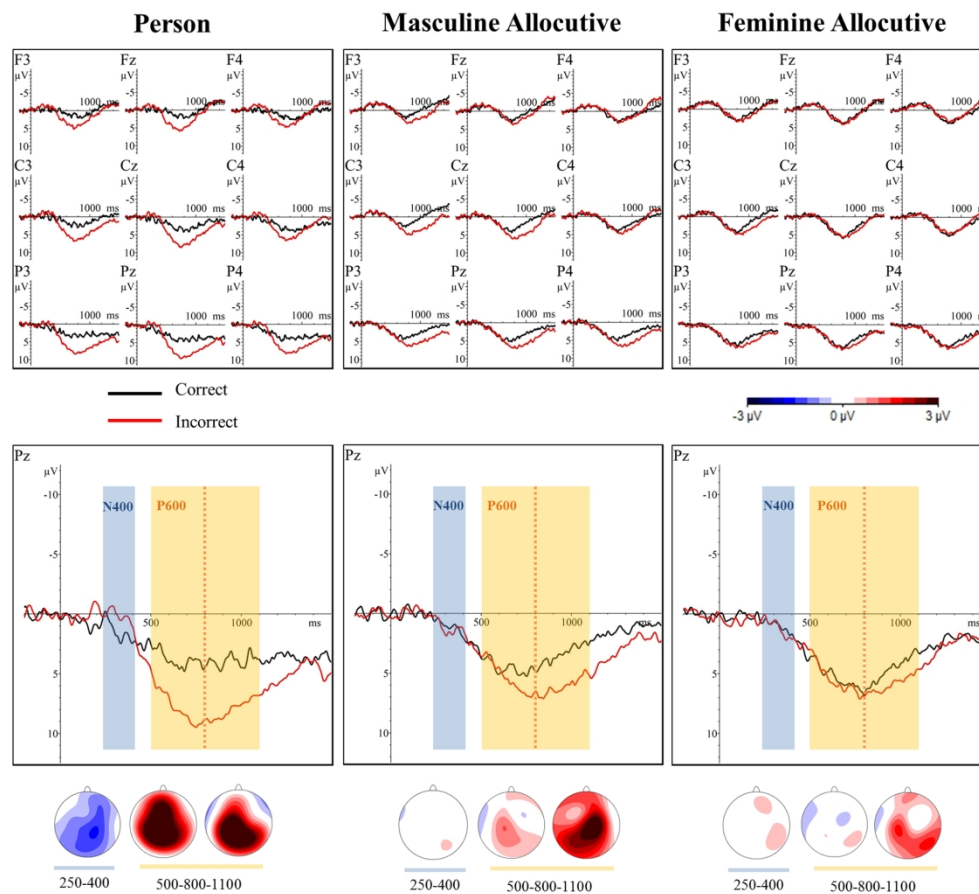
238x105mm (300 x 300 DPI)

Figure 3. Upper panel: ERP average waveforms for person, masculine and feminine allocutive from nine representative electrodes. Negativity is plotted upwards. Lower panel: ERP average waveforms are zoomed for the central posterior electrode (Pz). The time windows of interest are marked in color (N400: blue; P600: orange). Below each Pz plot, topographic distributions of ERP voltage differences (incorrect minus correct condition) are shown for each time window of interest.

189x181mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
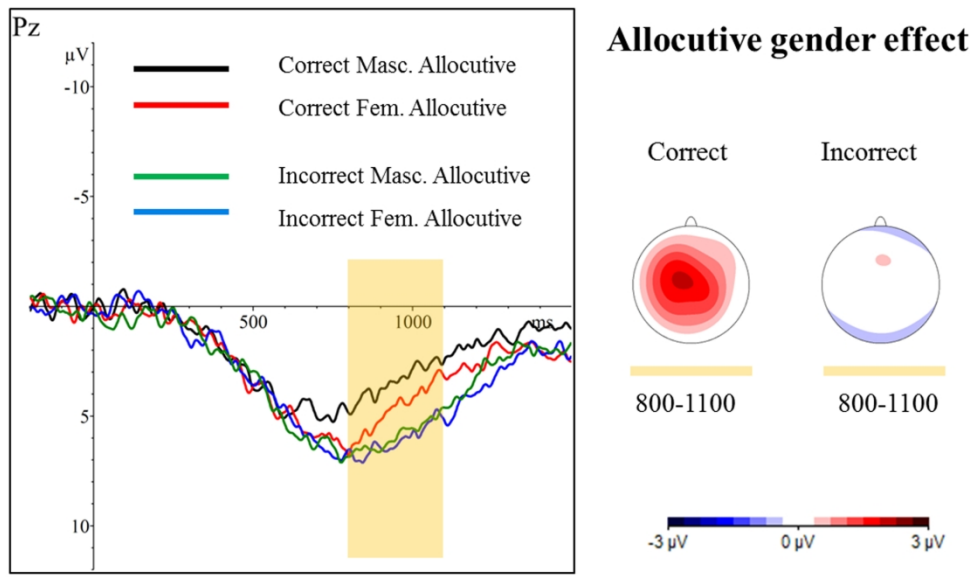43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 4. ERP average waveforms for correct and incorrect allocutive conditions are shown at Pz. On the right side, topographic distribution of Allocutive Gender effect (feminine minus masculine) are shown for the correct and the incorrect condition. The time window of interest (marked in color) is 800-1100 ms.

177x107mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
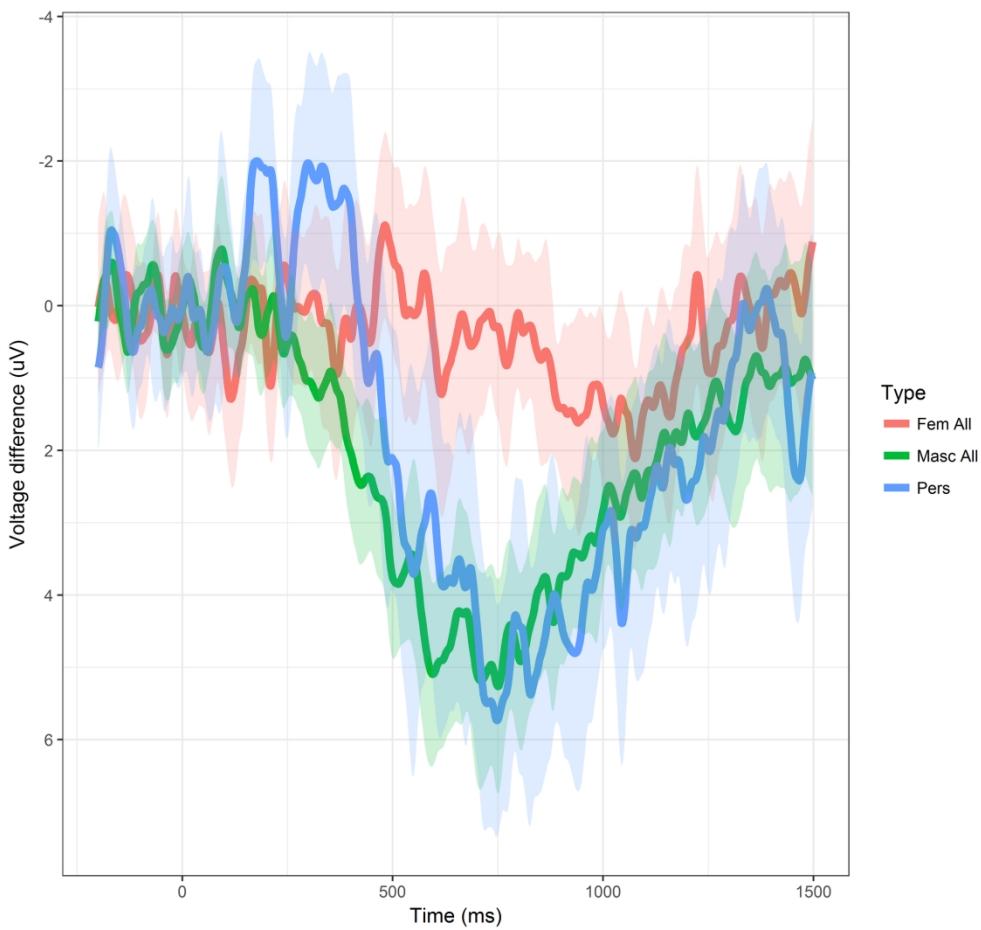51
52
53
54
55
56
57
58
59
60

Figure 5. ERP difference waveforms at Pz for feminine allocutive, masculine allocutive and person violation effects (incorrect minus correct). The shaded edges represent the 95% confidence intervals.

190x177mm (300 x 300 DPI)