

Eye gaze behaviour at turn transition: How aphasic patients process speakers' turns during video observation

Basil C. Preisig^{1,2}, Noëmi Eggenberger¹, Giuseppe Zito³, Tim Vanbellingen^{1,4}, Rahel Schumacher^{1,5}, Simone Hopfner¹, Klemens Gutbrod⁵, Thomas Nyffeler^{1,4}, Dario Cazzoli^{1,3}, Jean-Marie Annoni⁶, Stephan Bohlhalter^{1,4}, and René M. Müri^{1,3,5,7*}

¹Perception and Eye Movement Laboratory, Departments of Neurology and Clinical Research, Inselspital, University Hospital Bern, and University of Bern, Switzerland

²University Hospital of Old Age Psychiatry, University of Bern, Switzerland

³ARTORG Center for Biomedical Engineering Research, Gerontechnology and Rehabilitation Group, University of Bern, Switzerland

⁴Neurology and Neurorehabilitation Center, Department of Internal Medicine, Luzerner Kantonsspital, Switzerland

⁵Division of Cognitive and Restorative Neurology, Department of Neurology, Inselspital, University Hospital, and University of Bern, Switzerland

⁶Neurology Unit, Laboratory for Cognitive and Neurological Sciences, Department of Medicine, Faculty of Science, University of Fribourg, Switzerland

⁷Center for Cognition, Learning and Memory, University of Bern, Switzerland

* Address for correspondence: Prof. Dr. med. René Müri, Perception and Eye Movement Laboratory, Departments of Neurology and Clinical Research, Inselspital, University Hospital Bern, Freiburgstrasse 10, 3010 Bern, Switzerland. Phone: + 41 31 632 33 68; Fax: + 41 31 632 97 70; E-mail: Rene.mueri@insel.ch

Abstract

The human turn-taking system regulates the smooth and precise exchange of speaking turns during face-to-face interaction. Recent studies investigated the processing of on-going turns during conversation by measuring the eye movements of non-involved observers. The findings suggest that humans shift their gaze in anticipation to the next speaker before the start of the next turn. Moreover, there is evidence that the ability to timely detect turn transitions mainly relies on the lexico-syntactic content provided by the conversation. Consequently, patients with aphasia, who often experience deficits in both semantic and syntactic processing, might encounter difficulties to detect and timely shift their gaze at turn transitions. In order to test this assumption, we presented video vignettes of natural conversations to aphasic patients and healthy controls, while their eye movements were measured. The frequency and the latency of event-related gaze shifts, with respect to the end of the current turn in the videos, were compared between the two groups. Our results suggest that, compared to healthy controls, aphasic patients have a reduced probability to shift their gaze at turn transitions, but do not show significantly increased gaze shift latencies. In healthy controls, but not in aphasic patients, the probability to shift the gaze at turn transition was increased when the video content of the current turn had a higher lexico-syntactic complexity. Furthermore, the results from voxel-based lesion-symptom mapping indicate that the association between lexico-syntactic complexity and gaze shift latency in aphasic patients is predicted by brain lesions located in the posterior branch of the left arcuate fasciculus. Higher lexico-syntactic processing demands seem to lead to a reduced gaze shift probability in aphasic patients. This finding may represent missed opportunities for patients to place their contributions during everyday conversation.

Keywords: turn-taking, aphasia, turn projection, eye gaze, eye tracking, voxel-based lesion-symptom mapping

1 Introduction

The turn-taking system can be referred to as a speech exchange system, which organises the opportunities to speak during social interaction. Sacks, Schegloff, and Jefferson (1974) suggested that we are following a basic set of rules that are governing turn construction during conversation. For example, the current speaker has either the option to actively pass the turn to the next speaker (speaker's selection), or the turn can be taken by the listener at the next possible completion (self-selection). Following these basic rules ensures that there is only one speaker at a time. Apparently, self-selection requires that the listener is able to project the end of the turn. According to Sacks et al. (1974), this ability relies on our knowledge of the structure of the linguistic units, which enables us to project their ending in advance. As a consequence, this further allows us to project the end of a turn. This means that we recognise familiar linguistic units of a turn, and we are thus capable to project where the turn will end. At this point, one could ask how the turn-taking system, and with it turn projection, might be affected by a general disorder of language processing like aphasia. We approached this question by assessing eye movements from the perspective of non-involved observers, in order to evaluate the timing of gaze shifting at event-correlated turn transitions.

Aphasia is an acquired language disorder, and is a common consequence of brain damage to the language dominant hemisphere. The patients' impairments typically encompass both verbal production and verbal comprehension deficits which may alter conversational skills (Damasio, 1992). Nevertheless, previous research suggests that the fundamental communicative competence for effective turn-taking seems to be preserved in aphasic patients (Holland, 1982; Prinz, 1980; Schienberg & Holland, 1980; Ulatowska, Allard, Reyes, Ford, & Chapman, 1992). For instance, Schienberg and Holland (1982; 1980) reported, from the analysis of the conversation between two aphasic patients, that turn-taking behaviour remained intact. Aphasic patients even showed repair strategies for turn-taking errors when

both speakers were talking at the same time. The authors suggested that a naïve observer, who does not speak the language of the two patients, would not even notice their language production deficits. Even if turn-taking behaviour per se seems to be preserved, processing of linguistic information that has been shown to be crucial for the detection of turn transitions, seems to be impaired in aphasic patients. De Ruiter, Mitterer, and Enfield (2006) presented audio recordings from telephone conversations, which contained isolated turns. They found that healthy participants could reliably indicate the expected end of a turn before it was completed. The authors further reported that this ability depended on the availability of lexico-syntactic information. The intonational contour itself was not a sufficiently strong cue to anticipate the end of a turn. Consequently, aphasic patients, who often show deficits in semantic processing and / or syntactical processing (Caplan, Waters, Dede, Michaud, & Reddy, 2007; Caramazza & Berndt, 1978; Jefferies & Ralph, 2006), should also have greater difficulties to detect the linguistic units necessary in order to project the end of the turn.

Eye tracking has recently become a well-established technique to study the real-time processing of on-going turns in non-involved observers (Holler & Kendrick, 2015), but to the best of our knowledge has not been applied to aphasic patients. In this type of experimental paradigm, participants are requested to watch pre-recorded videos of dialogs, while their eye movements are recorded. The subsequent analysis focuses on the timing of participants' gaze shifts in relation to the turn transitions between speaking actors in the video. Studies using this paradigm consistently showed that participants track the current speaker with their eye gaze (Hirvenkari et al., 2013; Keitel & Daum, 2015; Keitel, Prinz, Friederici, von Hofsten, & Daum, 2013; Preisig et al., 2015; von Hofsten, Uhlig, Adell, & Kochukhova, 2009). Previous research suggests that the planning and execution of a saccadic gaze shifts require 200 ms (Becker, 1991; Griffin & Bock, 2000; Salthouse & Ellis, 1980; Westheimer, 1954). Hence, gaze shift reactions that occur within the first 200 ms after a turn is completed have been planned *a priori* and can thus be considered as indicators for turn end projection. Keitel et al.

(2013), as well as Keitel and Daum (2015) found that healthy individuals shift their gaze on the majority of turn transition in a time window spanning from 500 ms before the end of the current turn to the beginning of the next turn in the video. In contrast, Hirvenkari et al. (2013), who also presented pre-recorded video stimuli, did not find evidence for turn-related anticipatory gaze shifts in healthy participants. Holler and Kendrick (2015) explained the conflicting results by discrepancies in the stimulus properties, such as the different degree of spontaneity of the conversational exchange. Indeed, Hirvenkari et al. (2013) analysed only fast turn transitions in which the speaker change (from the end of one speaker's speech to the start of the other's) occurred within less than 300 ms. In contrast, turn transitions in the video material used by Keitel and colleagues lasted on average between 860 and 930 ms. This comparison suggests that anticipation of the next turn might be modulated by the duration of the inter-speaker gap. In order to exclude the impact of the inter-speaker gap duration, we decided to study turn projection as defined with respect to the end of the current turn.

In the present study, we addressed two main aspects of turn processing during video observation in aphasic patients: the detection of turn transitions, and the timing of transition related gaze shifts. We analysed the frequency of turn transition related gaze shifts as a measure of transition detection, and compared it with the gaze shift frequency on events without transition between speakers (i.e., pauses within a speaker's utterance or within-speaker overlaps). Pauses and within-speaker overlaps also have the potential to indicate a turn transition to the observer. We hypothesised that if aphasic patients have difficulties to detect turn transitions per se, they would also show less turn transition related gaze shifts. As a timing estimate, gaze shift latencies were calculated relative to the end of the current turn. In contrast to other studies, our video material was relatively fast-paced and not scripted. This led us to the expectation that the majority of gaze shifts would follow turn transitions, rather than precede them. Other than in previous studies, our stimulus material also included turn transitions with overlapping speech. Note that, at turn transitions with overlapping speech, the

next turn begins before the current turn ends. We distinguished between transitions with inter-speaker overlap and transitions with inter-speaker gap. Transitions with overlapping speech might be more difficult to project, because they happen suddenly, i.e., when the current speaker is interrupted by the next speaker (self-selection). Moreover, transitions with overlapping speech may represent a more ambiguous situation, in which it has to be resolved who is taking the next turn (Schegloff, 2000). The lexico-syntactic context helps healthy participants to reliably detect upcoming turn transitions (De Ruiter et al., 2006; Magyari & De Ruiter, 2012). Moreover, a current model of turn-taking assumes that humans rely on the linguistic content to make predictions about the unfolding of the current turn (Pickering & Garrod, 2013). However, content-rich sentences with increasing levels of lexico-syntactic complexity also impose higher processing demands on aphasic patients who are impaired in syntactic and / or semantic processing (Caplan et al., 2007; Jefferies & Ralph, 2006). Therefore, we expected that the advantage given by additional lexico-syntactic information would be only limited in aphasic patients. This led us to the hypothesis that higher lexico-syntactic complexity would either be related to a reduced detection of turn transitions, or to increased gaze shift latencies in aphasic patients. In a recent study, Keitel and Daum (2015) reported an additional gain of available intonation, as reflected in shorter turn transition related gaze shift latencies, in healthy participants. For this reason, we assessed whether the variance within the intonation curve, similar as the availability of intonation per se, would have an impact on turn processing. A higher gain triggered by the availability of video intonation would indicate a compensation of lexico-syntactic processing deficits in aphasic patients. Furthermore, we aimed to identify lesion sites associated with turn processing in aphasic patients, applying voxel-based lesion-symptom mapping (VLSM). VLSM is a method that allows to study the direct relationship between tissue damage and behaviour, on a voxel-by-voxel basis, comparable to functional neuroimaging (Bates et al., 2003).

2 Material & Methods

2.1 Participants

Sixteen aphasic patients with first-ever left-hemispheric stroke (mean age = 52.6, $SD \pm 13.3$ years; 5 females, 1 left-handed) and 23 healthy controls (mean age = 50.3, $SD \pm 16.4$ years; 8 females, 1 left-handed, 1 ambidexter) were included in the study. The present analysis is based on data that have been documented in a previous publication (Preisig et al., 2015). All participants had normal or corrected to normal visual acuity and an intact central visual field of 30°. At examination, patients were in a sub-acute to chronic condition (mean months post-stroke = 14.9, $SD \pm 16.3$). Aphasia diagnosis was based on standardised language assessment performed by clinical speech-language therapists. Aphasia severity was assessed by means of two subtests of the Aachen Aphasia Test (Huber, Poeck, & Willmes, 1984), namely the Token Test and the Written Language. Previous research demonstrated that the discriminative validity of these two subtests in conjunction is as good as the one of the whole test battery (Willmes, Poeck, Weniger, & Huber, 1980). Prior to study participation, written informed consent was obtained from all participants. The study was approved by the local Ethics Committee of the State of Bern and of the State of Luzern, and was conducted according to the latest version of the Declaration of Helsinki.

2.2 Procedure

Participants were seated in front of a SMI 250 Hz infrared eye-tracker (RED, SensoMotoric Instruments GmbH, Teltow, Germany), at a distance between 60 and 80 cm. After being seated, participants were instructed to attentively watch the presented videos. Prior to the main procedure, participants could familiarise with the experimental setting during a practice run. The four videos depicted dyadic conversations between a female and a male actor. The videos were presented on a 22" computer screen, and the actors in the video sequences covered a visual angle of approximately 16 degrees. Each video lasted 2 minutes. The order of

presentation was randomised. The experimental procedure lasted between 20 and 30 minutes. Further details concerning the video sequences and the procedure are provided in our recent report (Preisig et al., 2015).

2.3 Analysis of the video data

First, orthographic transcriptions of the video stimuli were time-aligned with the speech signal from the corresponding video audio file using the web service provided by the Bavarian Archive for Speech Signals (Kisler, Schiel, & Sloetjes, 2012; Schiel, 1999). The resulting TextGrid contained a time-aligned word segmentation of the speech signal. This TextGrid was then imported into the linguistic annotation software ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), where the time-alignment of the transcript was verified and manually adjusted if necessary.

For the transcribed video data, we defined the events that represented potential turn transition signals for the observer, and that could thus provoke a gaze shift away from the current speaker. According to Heldner and Edlund (2010), four event categories were defined: 1) overlap between speaker turns (inter-speaker overlap), 2) period of silence between speaker turns (inter-speaker gap), 3) period of silence within a speaker's utterance (pause), and 4) overlap within a speaker's utterance (within-speaker overlap) (see Fig. 1A). Even though only inter-speaker overlaps and inter-speaker gaps represent events with a turn transition, pauses and within-speaker overlaps can also elicit gaze shifts away from the current speaker (as indicated in Fig. 1B). The reason is that pauses and within-speaker overlaps can create an ambiguous situation where it is not clear for the observer who will take the next turn. The point of turn transition (i.e., turn relevance place) was set at the beginning of inter-speaker gaps and inter-speaker overlaps. For an overview of the video details (e.g., number and mean duration of events) see Table 1.

Figure 1

about here

In order to assess the content of each inter-event time interval (IETI), we calculated separate indices taking into account lexico-syntactic complexity and intonation. To ensure that enough lexico-syntactic and intonational information was provided during each IETI, events were included in the analysis only if they were preceded by an IETI that contained at least six words.

The lexico-syntactic complexity index was calculated as compound index, considering both the number and the median lexical frequency of the words during each IETI. We included separate measures for syntactic and lexical complexity because both properties can impose higher processing demands for patients with aphasia, and thus may affect their predictions of turn transitions. The number of words per IETI was taken as a measure of the syntactic load (Larsen-Freeman, 1978; Lu, 2011). Higher syntactic load requires higher phonological short-term memory capacities. Baldo and Dronkers (2006) found that aphasic patients show impairments in phonological short-term memory. The median lexical frequency was adopted as an indicator of lexical complexity, since more common words are usually correctly perceived at much lower speech to-noise ratios, a phenomenon referred to as the word frequency effect (Savin, 1963; Schuell, Jenkins, & Landis, 1961). Moreover, word frequency also affects lexical retrieval in aphasic patients (Luzzatti et al., 2002). The lexical frequency of the words within each IETI was calculated using WordGen (Duyck, Desmet, Verbeke, & Brysbaert, 2004). In order to build a compound index for lexico-syntactic complexity, lexical word frequency and syntactical complexity were combined using Stouffer's z-score method (Stouffer et al., 1949). Using this method, z-transformation was applied to the median word

frequency and to the number of words for each IETI. Please note that a lower word frequency corresponds to higher lexical complexity, and a higher number of words to higher syntactical complexity. Thus, the resulting z-scores were combined for each IETI by subtracting the z-standardized median word frequency from the z-standardized number of words. Subtraction of the z-scores takes into account that the combined values run into opposite directions. The values of the lexico-syntactic complexity index were also log-transformed with the natural logarithm, because their distribution was skewed.

As a measure of intonation, we considered the change in the intonation curve towards the end of the IETI. When a speaker's turn is coming to an end this can be indicated by a falling intonation, or by a rising intonation when asking a question (Bögels & Torreira, 2015; Duncan, 1972; Gravano & Hirschberg, 2011). Therefore, we decided to take the variance in the intonation curve during the last six words of the IETI as prosodic turn signal. For this purpose, the base frequency (f_0) of the video sound files in Hz was extracted using the Praat software (Boersma & Weenink, 2001). Then, the variance within f_0 was calculated over the last six words of each IETI.

2.4 Analysis of eye movement data

Saccadic data were extracted from the SMI analysis software (BeGazeTM, SensoMotoric Instruments GmbH, Teltow, Germany). Only direct gaze shifts between the face regions of the two actors in the video were included in the analysis, i.e., saccades that started on the face region of one actor and ended on the face region of the other actor.

Event-related gaze shifts were selected for the analysis by means of a crucial time window. Every saccadic gaze shift occurring in a time window ranging from 1000 ms before to 1000 ms after an event was included in the analysis. Events were considered for analysis only if the preceding and the subsequent IETIs lasted at least 1000 ms. The aim of this procedure was to

prevent that the crucial time window of one given event would overlap with the one of another event occurring right beforehand or just afterwards.

Furthermore, we only included in the analysis: a) gaze shifts in the direction of the corresponding turn transition, i.e., from the current to the next speaker; and, b) in case of an event without transition (pauses and within-speaker overlaps), gaze shifts leading away from the current speaker. Thus, random gaze shifts directed from the listener to the speaker were not included.

Two dependent variables were computed for every event that was included in the analysis: the binomial variable gaze shift reaction (0 = no gaze shift; 1 = gaze shift), and the continuous variable gaze shift latency in milliseconds. If a participant produced multiple gaze shifts within the crucial time window of a single event, only the first gaze shift was considered as relevant for the analysis. The gaze shift latency was calculated by subtracting the starting time of the saccade from the starting time of the corresponding event. Thus, a negative value indicates that the starting time of the saccade preceded the starting time of the event, and vice versa. The average gaze shift frequency per participant was calculated as the ratio: number of gaze shifts per event category, divided by the number of events per category.

Table 1
about here

2.5 Data analysis

Statistical analyses were conducted with IBM Statistics SPSS 21 and lme4 (Bates, Mächler, Bolker, & Walker, 2014), a package implemented in the open-source program R (R Core Team, 2014). Two separate repeated-measures analyses of variance (ANOVAs) were calculated for the dependent variables average gaze shift frequency and average gaze shift

latency. For post-hoc comparisons, pairwise t-tests with Holm correction were calculated. Partial eta squared was computed as an estimate of effect size.

In order to take into account variables that unfolded during the course of our experiment (such as lexico-syntactic complexity and intonation variance during the IETI), we applied mixed effect modelling using the lme4 package. A key advantage of mixed-effects models is that they do not require prior averaging (Baayen, Davidson, & Bates, 2008), since each participant has its own intercept, which randomly deviates from the mean intercept. Therefore, individual gaze shift reactions, which occurred in relation to different events, can be directly entered into the model. A generalized linear mixed model (GLMM) for binomial data was calculated for the dependent variable gaze shift reaction (0 = no gaze shift; 1 = gaze shift) using the glmer function. The glmer function provides *p*-values for the fixed effects in the model based on asymptotic Wald tests. Least-square means were computed for post-hoc comparisons in the GLMM. Furthermore, a linear mixed model (LMM) was calculated for the continuous variable gaze shift latency, applying the lmer function. For this model, the analysed data were unbalanced, because participants shifted their gaze in relation to different events in the videos. Therefore, the lmer function cannot apply simple formulas in order to estimate the degrees of freedom. For this reason, in such cases the lmer function provides only a list of *t*-values, but no *p*-values. However, when the number of subjects and the number of observations is sufficiently large, there is a strong correspondence between the *t*-statistics and the *z*-statistics. In this case, *t*-values larger than ± 2 can be considered as significant (Baayen et al., 2008; Ohi, Brandt, & Kliegl, 2011).

2.6 Lesion mapping

Lesion analysis of imaging data was conducted using the open source software MRICron (Rorden, Karnath, & Bonilha, 2007). The brain lesions of 11 patients with available MRI scans (volume of interest (VOI)) were delineated directly onto the transversal slices of the

individual T2-weighted MRI scans. The MRI scan of each patient, including the lesion VOI, was then normalised into the Talairach space using the spatial normalisation algorithm provided by SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/>). The brain lesions of the five remaining patients, with an available computed tomography (CT) scan, were mapped directly onto the CH2 template brain implemented in MRICron (Rorden & Brett, 2000). In order to relate behavioural measures to neuroanatomy, conventional lesion subtraction and voxel-based lesion-symptom mapping (VLSM) analyses were conducted. For the lesion subtraction analysis, which only provides descriptive outcomes, lesion VOIs of patients who showed positive correlations between gaze shift latencies with lexico-syntactic complexity and intonation variance were contrasted with lesion VOIs of patients who showed the opposite pattern. MRICron offers two VLSM methods: the nonparametric Lieberman Test for binomial data, and *t*-tests for continuous behavioural data. We applied both methods, aiming to find converging evidence through these two types of analysis. Only voxels surviving a conservative permutation thresholding with family-wise error (FWE-corrected level of $p < .01$) correction were considered in the results. Furthermore, voxels that were damaged in less than 20% of the patients were excluded from the analysis.

3 Results

3.1 Average gaze shift frequency

For the average frequency of gaze shifts per event category, a three-way repeated measures ANOVA with turn transition (transition; no transition), and event type (overlapping speech; silence) as within-subject factors, and group (aphasic patients; healthy controls) as between-subject factor revealed a significant main effect of the factor turn transition [$F(1,37) = 129.741, p < .001, \eta_p^2 = .778$], and a significant interaction between factors turn transition x group [$F(1,37) = 5.541, p = .024, \eta_p^2 = .130$]. As expected, the frequency of gaze shifts depended on the factor turn transition. Participants were more likely to react to turn transitions

as compared to events without transition (pauses and within-speaker overlaps, see Fig. 2). More interestingly, a post hoc comparison on the average gaze shift frequency at turn transition revealed a statistical trend towards a group difference ($p = .072$). Aphasic patients tended to show a lower average gaze shift frequency at turn transition than healthy controls. The analysis of the gaze shift frequency per event category demonstrated that turn transitions elicited more gaze shifts than events without transition ($p < .001$). Hence, the subsequent analyses only focused on the processing of turn transitions.

Figure 2

about here

3.2 Gaze shift probability at turn transitions

A GLMM for the binomial data was modelled for the dependent variable gaze shift reaction (0 = no gaze shift; 1 = gaze shift) including the fixed factors group (aphasic patients; healthy controls), type of turn transition (inter-speaker gap, inter-speaker overlap), lexico-syntactic complexity, and intonation variance. Furthermore, subject and video were controlled as random effect terms.

The GLMM revealed significant main effects of group ($z = -1.988, p = .047$), type of turn transition ($z = -3.167, p = .002$), and lexico-syntactic complexity ($z = -3.529, p < .001$). Healthy controls showed a higher probability than aphasic patients to shift their gaze on turn transitions. The probability for a gaze shift reaction was higher for turn transitions with inter-speaker gap, and for turn transitions that were preceded by an IETI with a higher lexico-syntactic complexity. Moreover, the GLMM revealed significant interactions between factors group x lexico-syntactic complexity ($z = 2.847, p = .004$), group x type of turn transition ($z = 3.243, p < .001$), lexico-syntactic complexity x type of turn transition ($z = 3.429, p < .001$),

and a trend towards an interaction between factors group x type of turn transition x lexico-syntactic complexity ($z = -1.889, p = .059$). In healthy participants, the probability for a gaze shift increased with increasing values on the lexico-syntactic complexity index, while the opposite pattern could be observed in aphasic patients (see Fig. 3a). Furthermore, healthy controls showed a higher probability for a gaze shift reaction on inter-speaker gaps compared to inter-speaker overlaps ($p = .049$) (see Fig. 3b). The interaction between factors lexico-syntactic complexity x type of turn transition revealed two opposing patterns: For inter-speaker overlaps gaze shift probability decreased with increasing lexico-syntactic complexity, for inter-speaker gaps we observed the reversed pattern. The statistical trend for the three-way interaction between factors group x type of turn transition x lexico-syntactic complexity suggests that aphasic patients reacted less frequently to inter-speaker overlaps with higher lexico-syntactic complexity than healthy participants. There was neither a main effect of factor intonation variance ($z = -0.782, p = .434$), nor an interaction between factors group x intonation variance ($z = 0.542, p = .588$).

3.3 Gaze shift latency at turn transition

Gaze shifts followed, rather than preceded, the turn transitions in the video ($M_{\text{Patients/Inter-speaker gap}} = 280.82, SE_{\text{Patients/Inter-speaker gap}} = 69.11; M_{\text{Patients/Inter-speaker overlap}} = 248.21; SE_{\text{Patients/Inter-speaker overlap}} = 79.48, M_{\text{Controls/Inter-speaker gap}} = 226.36, SE_{\text{Controls/Inter-speaker gap}} = 44.24, M_{\text{Controls/Inter-speaker overlap}} = 158.81, SE_{\text{Controls/Inter-speaker overlap}} = 66.63$). The repeated measures ANOVA on the dependent variable average gaze shift latency did neither reveal significant main effects for group or type of turn transition, nor an interaction between these two factors. Furthermore, a LMM was modelled on the dependent variable gaze shift latency at turn transition. The LMM included the same fixed factors and random factors as the GLMM introduced above. We found significant interactions between factors group x intonation variance ($t = 2.116$) and group x lexico-syntactic complexity x intonation variance ($t = -2.249$). As depicted in Figure

4, healthy participants showed shortest gaze shift latencies if both lexico-syntactic complexity and intonation variance were increased. In contrast, aphasic patients did not show such a clear pattern.

Figure 3+4

about here

3.4 Lesion analysis

The mean volume of aphasic patients' individual brain lesions was 96.14 cm³ (*SD* = 17.00 cm³). One patient was excluded from the lesion analysis because he was left-handed. In order to identify lesion sites associated with increased gaze shift latencies due to processing of lexico-syntactic and intonation variance, we performed two separate VLSM analyses: one with two binary predictors, and another with two continuous predictors. For this reason, correlations coefficients were calculated for each participant between the lexico-syntactic complexity and the intonation variance during the IETI, with the gaze shift latency registered at the corresponding turn transition. A positive correlation indicates that higher lexico-syntactic complexity and / or more intonation variance is associated with increased gaze shift latencies. In contrast, a negative correlation indicates that a participant can benefit from additional lexico-syntactic content or more intonation variance, as reflected in shorter gaze shift latencies. For the first VLSM model, correlation coefficients for the variables lexico-syntactic complexity and intonation variance were transformed into separate binary predictors. To binarize the predictor, the 75th percentile was used as cut-off score, since higher positive correlations are related to maladaptive processing (i.e., increased gaze shift latencies). This resulted in the following two values: 0 (correlation coefficient > 75th

percentile of the correlation coefficients obtained in healthy participants), and 1 (correlation coefficient < 75th percentile of the correlation coefficients obtained in healthy participants).

The binomial Liebermeister Test, calculated for the first VLSM model, revealed a significant lesion cluster (FWE-corrected level of $p < .01$) for the factor lexico-syntactic complexity on the posterior branch of the left arcuate fasciculus (Talairach coordinates; -37, -46, 24), but no significant cluster for the factor intonation variance (Fig. 5). This result was confirmed by the second VLSM model, where the individual correlation coefficients were entered as a continuous predictor and t -tests (FWE-corrected level of $p < .01$) were applied to perform comparisons on a voxel-by-voxel basis (Talairach coordinates; -36, -48, 24). Furthermore, we verified the reliability of the VLSM models by running an additional lesion subtraction analysis. Distinct lesion overlap maps were generated for the two patient subgroups defined according to the factor lexico-syntactic complexity. In line with the VLSM analyses, the group subtraction analysis revealed that patients with increased gaze shift latencies at turn transitions, due to higher lexico-syntactic complexity of the preceding turn, showed an exclusive lesion cluster on the posterior branch of the arcuate fasciculus (Talairach coordinates; -37,-47, 23) (Fig. 6C).

Figure 5+6

about here

4 Discussion

The present study aimed at gaining new insights into real-time processing of on-going turns in aphasic patients by analysing the frequency and the timing of turn transition related gaze shifts during video observation of naturalistic conversations. The main finding is that aphasic patients showed a lower probability to shift their gaze at turn transitions than healthy

participants. The probability whether a gaze shift would occur or not depended on the lexico-syntactic complexity of the video content preceding a particular turn transition. In healthy controls, higher lexico-syntactic complexity led to higher gaze shift probabilities. The opposite, i.e. decreasing gaze shift probability associated with higher lexico-syntactic complexity, was found in aphasic patients. The timing of gaze shifts depended on both the lexico-syntactic complexity and on the intonation variance provided before turn transitions. Healthy controls, but not aphasic patients, gained from intonation variance when lexico-syntactic complexity was increased. Furthermore, we found that brain lesions to the posterior branch of the left arcuate fasciculus predicted the impact of lexico-syntactic complexity on gaze shift latency in aphasic patients.

Our results indicate that turn transitions trigger more gaze shifts in both groups than pauses and within-speaker overlaps. This implies that aphasic patients did not show unsystematic visual exploration behaviour during video observation, and that they were per se able to reliably detect turn transitions. However, aphasic patients showed a lower probability to react to turn transitions than healthy controls. This observation is supported by converging evidence from the repeated measures ANOVA conducted on average gaze shift frequency over all event categories, and from the GLMM including the binomial data from individual turn transitions in the video material.

Overall, gaze shift probability at turn transitions depended on the type of turn transition, and on the complexity of the lexico-syntactic information provided before the transition itself, but not on the intonation variance in the same time window. This finding fit well with evidence from previous research, which suggested that the ability to detect upcoming turn transitions mainly relies on the availability of lexico-syntactic information (De Ruiter et al., 2006). Our results imply that additional lexico-syntactic information may help healthy participants to detect upcoming turn transitions more accurately. Gaze shift probability in healthy participants was higher for transitions with inter-speaker gaps. This supports our hypothesis that this type

of turn transition is more reliably detected. In the case of an inter-speaker gap, it is probably easier to resolve who is taking the next turn than for inter-speaker overlaps (Schegloff, 2000). Moreover, healthy participants shifted their gaze more frequently on turn transitions that were preceded by segments with higher lexico-syntactic complexity. As expected, lexico-syntactic complexity had an opposite effect on gaze shift probability in aphasic patients suggesting that aphasic patients have difficulties to integrate this parameter when initiating their gaze-shift. Previous research clearly indicates impairments in lexico-syntactic processing in aphasic patients (Caplan et al., 2007; Caramazza & Berndt, 1978; Jefferies & Ralph, 2006). Moreover, in a recent eye-tracking study from our group, we found reduced understanding of syntactically complex sentences in aphasic patients due to impaired recognition and integration of morphosyntactic cues (Schumacher et al., 2015). Therefore, the video segments with higher lexico-syntactic complexity might be very demanding for aphasic patients, and were thus accompanied by a reduced gaze shift probability.

In line with the results reported by Hirvenkari et al. (2013), we found that the majority of gaze shifts occurred after turn transitions. The average gaze shift latency indicates that some gaze shifts were planned prior to the completion of the turn. Contrary to our expectation, aphasic patients did not show an increased gaze shift latency. However, we found an interaction between group membership (aphasic patients; healthy controls), lexico-syntactic complexity, and intonation variance. Healthy controls were most likely to project the end of a speaker turn when lexico-syntactic complexity and intonation variance were increased (Fig. 4). In a recent study, Keitel and Daum (2015) also found an effect of intonation on gaze shift latencies in healthy participants. According to our results, aphasic patients did not benefit from the interaction between intonation variance and lexico-syntactic complexity. Earlier studies found that the processing of linguistic prosody mainly relies on the right hemisphere (Brådvik et al., 1991; Weintraub, Mesulam, & Kramer, 1981) which would suggest that the recognition of linguistic prosody should not be affected in aphasic patients with left hemispheric brain

lesions. This finding has later been supported by neuroimaging studies in healthy subjects which reported right hemispheric specialization when linguistic prosody was compared with other speech processes (Kyong et al., 2014; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Strelnikov, Vorobyev, Chernigovskaya, & Medvedev, 2006). In contrast, studies which directly compared linguistic prosody with emotional prosody found a primary involvement of the left brain hemisphere (Pell & Baum, 1997; Wildgruber et al., 2004). Shapiro and Nagel (1995) suggested that aphasic patients with deficits in lexico-syntactic processing may not benefit from additional prosodic information when parsing sentential units because they cannot concurrently process syntactic and prosodic information. This notion is supported by a recent imaging study in healthy participants, which suggests that the recognition of linguistic prosody depends on the activity in a bilateral network (Kreitewolf, Friederici, & von Kriegstein, 2014).

Voxel-based lesion-symptom mapping revealed that the modulation of the gaze shift latencies by the lexico-syntactic complexity was predicted by a lesion cluster located on the posterior end of the arcuate fasciculus, between the inferior parietal lobe and the superior temporal lobe. This area is part of the left-hemispheric language network, lying in close vicinity of the left posterior superior temporal gyrus, the left angular gyrus, and the temporo-parietal junction. Functional imaging studies in healthy participants showed that both the left superior temporal gyrus and the left angular gyrus are involved in syntactic and semantic processing (Buchsbaum & D'Esposito, 2008; Buchsbaum, Hickok, & Humphries, 2001; Graves, Grabowski, Mehta, & Gupta, 2008; Humphries, Binder, Medler, & Liebenthal, 2006; Keller, Carpenter, & Just, 2001).

Several studies consistently reported that the left posterior superior temporal cortex is activated during written sentence comprehension (Cooke et al., 2002; Just, Carpenter, Keller, Eddy, & Thulborn, 1996) and auditory sentence comprehension (Buchsbaum et al., 2001). Moreover, activity in the left superior temporal gyrus seems to be modulated by syntactic

complexity (Cooke et al., 2002; Friederici, Makuuchi, & Bahlmann, 2009; Just et al., 1996; Kinno, Kawamura, Shioda, & Sakai, 2008; Newman, Ikuta, & Burns, 2010) and word frequency (Graves et al., 2008). However, this area seems not only to be involved in language processing, but also in audio-visual integration (Beauchamp, Lee, Argall, & Martin, 2004; Calvert, Campbell, & Brammer, 2000; Stevenson & James, 2009) and face processing (Haxby, Petit, Ungerleider, & Courtney, 2000). Taken together, the left superior temporal cortex is clearly involved in integrating different types of information during language processing. Friederici (2011) suggested in a review article that the left posterior superior temporal cortex together with the superior temporal sulcus and the basal ganglia might be involved in the integration of semantic and syntactic information.

The left angular gyrus also seems to support both sentence-level semantic and syntactic processing. Bavelier et al. (1997) showed increased activation of the left angular gyrus in response to additional syntactic information, i.e., when comparing sentence reading with word list reading. Humphries et al. (2006) found greater activity in the left angular gyrus for semantically congruent sentences compared to sentences containing random words or pseudowords. Humphries and colleagues (2006) further suggested that the left angular gyrus might be more strongly engaged in semantic processes than in syntactic ones, because it requires lexical information to be activated, but not necessarily syntactic information. Interestingly, Keller et al. (2001) found that angular gyrus activation interacts with lexical word frequency, showing stronger activation for more complex sentences that include low-frequency words than less complex sentences.

The left temporo-parietal junction (TPJ), also known as left sylvian–parietal–temporal area, is thought to function as a sensorimotor interface between the phonological networks in the bilateral superior temporal gyrus and the articulatory networks in the anterior language system (Hickok & Poeppel, 2007). This area has also been shown to be crucial for auditory verbal working memory (Buchsbaum & D'Esposito, 2008; Shallice & Warrington, 1977).

Furthermore, the left TPJ, which is located laterally with respect to the lesion cluster identified in the present study, also belongs to a widely distributed neural network underlying Theory of Mind mechanisms. Ciaramidaro et al. (2007) investigated the contribution of different nodes within this network in an fMRI study showing that left TPJ was selectively activated when participants had to anticipate the endings of stories that included social and especially communicative intentions.

Conclusion

In the present study, we investigated the detection of turn transitions and the timing of transition related gaze shifts in aphasic patients. We found that the detection of turn transitions in healthy participants depends on the lexico-syntactic information provided before the transition itself. Moreover, healthy controls were more likely to project the end of turns when higher lexico-syntactic complexity was associated with a greater amount of prosodic information. We showed that gaze shift probability in aphasic patients was reduced at transitions that were preceded by more complex turns with higher lexico-syntactic processing demands.

5 Acknowledgments

This study was supported by the Swiss National Science Foundation [Grant no. 320030_138532/1]. We also would like to thank Sandra Perny, Susanne Zürcher, Julia Renggli, Marianne Tschirren, Corina Wyss, Carmen Schmid, Gabriella Steiner, Monica Koenig-Bruhin, Nicole Williams, Reto Hänni, Gianni Pauciello, Silvia Burren, Andreas Löffel, Michael Schraner, Nina Kohler, Anita Mani-Luginbühl, Hans Witschi, Sarah Schaefer, Martin Zürcher, and Michael Rath for their assistance.

6 References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.
- Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, *20*(5), 529-538.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. Retrieved from: <http://arxiv.org/abs/1406.5823>
- Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*, *6*(5), 448-450.
- Bavelier, D., Corina, D., Jezzard, P., Padmanabhan, S., Clark, V., Karni, A., . . . Rauschecker, J. P. (1997). Sentence reading: A functional MRI study at 4 Tesla. *Journal of Cognitive Neuroscience*, *9*(5), 664-686.
- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809-823.
- Becker, W. (1991). Saccades in Vision and Visual Dysfunction . In R. H. S. Carpenter (Ed.), *Eye Movements* (pp. 95–137). London: Macmillan.
- Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer [Computer program]. Version 5.4.18. Retrieved from: <http://www.praat.org>
- Brådvik, B., Dravins, C., Holtås, S., Rosen, I., Ryding, E., & Ingvar, D. (1991). Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurologica Scandinavica*, *84*(2), 114-126.
- Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience*, *20*(5), 762-778.
- Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*(5), 663-678.
- Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46-57.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649-657.
- Caplan, D., Waters, G., Dede, G., Michaud, J., & Reddy, A. (2007). A study of syntactic processing in aphasia I: behavioral (psycholinguistic) aspects. *Brain and Language*, *101*(2), 103-150.

- Caramazza, A., & Berndt, R. S. (1978). Semantic and syntactic processes in aphasia – review of literature. *Psychological Bulletin*, 85(4), 898-918.
- Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., & Walter, H. (2007). The intentional network: how the brain reads varieties of intentions. *Neuropsychologia*, 45(13), 3105-3113.
- Cooke, A., Zurif, E. B., DeVita, C., Alsop, D., Koenig, P., Detre, J., . . . Grossman, M. (2002). Neural basis for sentence comprehension: grammatical and short-term memory components. *Human Brain Mapping*, 15(2), 80-94.
- Damasio, A. R. (1992). Aphasia. *The New England Journal of Medicine*, 326(8), 531-539.
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), 515-535.
- De Schotten, M. T., Bizzi, A., Dell'Acqua, F., Allin, M., Walshe, M., Murray, R., . . . Catani, M. (2011). Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *Neuroimage*, 54(1), 49-59.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2), 283-292.
- Duyck, W., Desmet, T., Verbeke, L. P., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers*, 36(3), 488-499.
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews*, 91(4), 1357-1392.
- Friederici, A. D., Makuuchi, M., & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *Neuroreport*, 20(6), 563-568.
- Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3), 601-634.
- Graves, W. W., Grabowski, T. J., Mehta, S., & Gupta, P. (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of Cognitive Neuroscience*, 20(9), 1698-1710.
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 11(4), 274-279.
- Haxby, J. V., Petit, L., Ungerleider, L. G., & Courtney, S. M. (2000). Distinguishing the functional roles of multiple regions in distributed neural systems for visual working memory. *Neuroimage*, 11(2), 145-156.
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555-568.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.

- Hirvenkari, L., Ruusuvoori, J., Saarinen, V. M., Kivioja, M., Perakyla, A., & Hari, R. (2013). Influence of turn-taking in a two-person conversation on the gaze of a viewer. *Plos One*, 8(8). Retrieved from <http://www.plosone.org>
- Holland, A. L. (1982). Observing functional communication of aphasic adults. *Journal of Speech and Hearing Disorders*, 47(1), 50-56.
- Holler, J., & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing reciprocity. *Frontiers in Psychology*, 6(14). Retrieved from <http://www.frontiersin.org>
- Huber, W., Poeck, K., & Willmes, K. (1984). The Aachen Aphasia Test. *Advances in Neurology*, 42, 291-303.
- Humphries, C., Binder, J. R., Medler, D., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18(4), 665-679.
- Jefferies, E., & Ralph, M. A. L. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain*, 129, 2132-2147.
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274(5284), 114-116.
- Keitel, A., & Daum, M. M. (2015). The use of intonation for turn anticipation in observed conversations without visual signals as source of information. *Frontiers in Psychology*, 6. Retrieved from <http://www.frontiersin.org>
- Keitel, A., Prinz, W., Friederici, A. D., von Hofsten, C., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology*, 116(2), 264-277.
- Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. *Cerebral cortex*, 11(3), 223-237.
- Kinno, R., Kawamura, M., Shioda, S., & Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture - sentence matching task. *Human brain mapping*, 29(9), 1015-1027.
- Kisler, T., Schiel, F., & Sloetjes, H. (2012). Signal processing via web services: the use case WebMAUS. In *Proceedings of the Digital Humanities 2012*, Hamburg, Germany.
- Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *Neuroimage*, 102, 332-344.
- Kyong, J. S., Scott, S. K., Rosen, S., Howe, T. B., Agnew, Z. K., & McGettigan, C. (2014). Exploring the roles of spectral detail and intonation contour in speech intelligibility: An fMRI study. *Journal of Cognitive Neuroscience*, 26(8), 1748-1763.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 439-448.

- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers language development. *TESOL Quarterly*, 45(1), 36-62.
- Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G.-D. (2002). Verb-noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and Language*, 81(1), 432-444.
- Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3. Retrieved from <http://www.frontiersin.org>
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human brain mapping*, 17(2), 73-88.
- Newman, S. D., Ikuta, T., & Burns, T. (2010). The effect of semantic relatedness on syntactic analysis: an fMRI study. *Brain and Language*, 113(2), 51-58.
- Ohl, S., Brandt, S. A., & Kliegl, R. (2011). Secondary (micro-) saccades: The influence of primary saccade end point and target eccentricity on the process of postsaccadic fixation. *Vision research*, 51(23), 2340-2347.
- Pell, M. D., & Baum, S. R. (1997). Unilateral brain damage, prosodic comprehension deficits, and the acoustic cues to prosody. *Brain and Language*, 57(2), 195-214.
- Pickering, M. J., & Garrod, S. (2013). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, 36(4), 377-392.
- Preisig, B. C., Eggenberger, N., Zito, G., Vanbellingen, T., Schumacher, R., Hopfner, S., . . . Müri, R. M. (2015). Perception of co-speech gestures in aphasic patients: A visual exploration study during the observation of dyadic conversations. *Cortex*, 64, 157-168.
- Prinz, P. M. (1980). A note on requesting strategies in adult aphasics. *Journal of Communication Disorders*, 13(1), 65-73.
- Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, 12(4), 191-200.
- Rorden, C., Karnath, H. O., & Bonilha, L. (2007). Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience*, 19(7), 1081-1088.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.
- Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *The American Journal of Psychology*, 207-234.
- Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *The Journal of the Acoustical Society of America*, 35(2), 200-206.
- Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(01), 1-63.

- Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. *In Proceedings of the 14 ICPHS*, San Francisco, USA.
- Schienenberg, S., & Holland, A. (1980). Conversational turn-taking in wernike aphasia. In R. H. Brookshire (Ed.), *Proceedings of the Clinical Aphasiology Conference*, BRK Publishers, Minneapolis, Minnesota, USA.
- Schuell, H., Jenkins, J., & Landis, L. (1961). Relationship between auditory comprehension and word frequency in aphasia. *Journal of Speech, Language, and Hearing Research*, 4(1), 30-36.
- Schumacher, R., Cazzoli, D., Eggenberger, N., Preisig, B., Nef, T., Nyffeler, T., . . . Müri, R. M. (2015). Cue recognition and integration: eye tracking evidence of processing differences in sentence comprehension in Aphasia. *PloS one*, 10(11), Retrieved from <http://www.plosone.org>
- Shallice, T., & Warrington, E. K. (1977). Auditory-verbal short-term memory impairment and conduction aphasia. *Brain and Language*, 4(4), 479-491.
- Shapiro, L. P., & Nagel, H. N. (1995). Lexical properties, prosody, and syntax: implications for normal and disordered language. *Brain and Language*, 50(2), 240-257.
- Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*, 44(3), 1210-1223.
- Stouffer, S. A., Lumsdaine, A. A., Lumsdaine, M., Williams, R. M., Smith, M. B., Janis, I. L. ... Cottrell, L. S. (1949). *The American Soldier: Adjustment during Army Life*. Princeton: Princeton University Press.
- Strelnikov, K. N., Vorobyev, V. A., Chernigovskaya, T. V., & Medvedev, S. V. (2006). Prosodic clues to syntactic processing: a PET and ERP study. *Neuroimage*, 29(4), 1127-1134.
- Ulatowska, H. K., Allard, L., Reyes, B. A., Ford, J., & Chapman, S. (1992). Conversational discourse in aphasia. *Aphasiology*, 6(3), 325-330.
- von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders*, 3(2), 556-569.
- Weintraub, S., Mesulam, M. M., & Kramer, L. (1981). Disturbances in prosody: a right-hemisphere contribution to language. *Archives of Neurology*, 38(12), 742-744.
- Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *AMA Archives of Ophthalmology*, 52(6), 932-941.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, 14(12), 1384-1389.
- Willmes, K., Poeck, K., Weniger, D., & Huber, W. (1980). The aachener aphasia test – differential validity. *Nervenarzt*, 51(9), 553-560.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. *In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Nijmegen, the Netherlands.

7 Figure Captions

Figure 1

(A) The Illustration of the different event categories: inter-speaker overlap, inter-speaker gap, pause, and within-speaker overlap. The red lines indicate the point of turn transition. (B) The distribution of individual gaze shifts (each grey dot represents the start of a single saccade) over a time window of one minute. The grey line represents a density estimate of the saccade distribution. The colours of the bars correspond to the different event categories introduced in (A).

Figure 2

Average frequency of gaze shifts per event category illustrating the main effect of the factor turn transition. Asterisks indicate the level of significance (* $p < .001$)

Figure 3

(A) Gaze shift probability at turn transitions as a function of lexico-syntactic complexity in aphasic patients and healthy participants. Density rugs illustrate gaze shift probability at different levels of lexico-syntactic complexity. (B) Gaze shift probability as a function of turn transition type (inter-speaker overlap, inter-speaker gap).

Figure 4

Gaze shift latency as a function of lexico-syntactic complexity and prosodic information in aphasic patients and healthy participants. In the 3D scatterplot with regression plane the red dots illustrate individual gaze shift latencies at different levels of lexico-syntactic complexity and prosodic information.

Figure 5

Results of the voxel-based lesion symptom mapping. Voxels with damage that was a significant predictor for increased gaze shift latencies due to higher lexico-syntactic complexity in purple. The representation of the entire left arcuate fasciculus (in blue) is based

on a recently published DTI atlas (De Schotten et al., 2011). Talairach coordinates are presented at the bottom of the figure.

Figure 6

A and B) Overlap maps of the brain lesions in the two patient subgroups (A: patients benefiting from lexico-syntactic complexity; B: patients impaired by lexico-syntactic complexity). C) Results of the lesion subtraction analysis. 50 to 100% of the patients belonging to subgroup B (patients impaired by lexico-syntactic complexity) have a brain lesion in this area, but not patients belonging to subgroup A (patients benefiting from lexico-syntactic complexity). The z-position of each axial slice in the Talairach stereotaxic space is presented at the bottom of the figure.

Figure 1. Definition of event categories

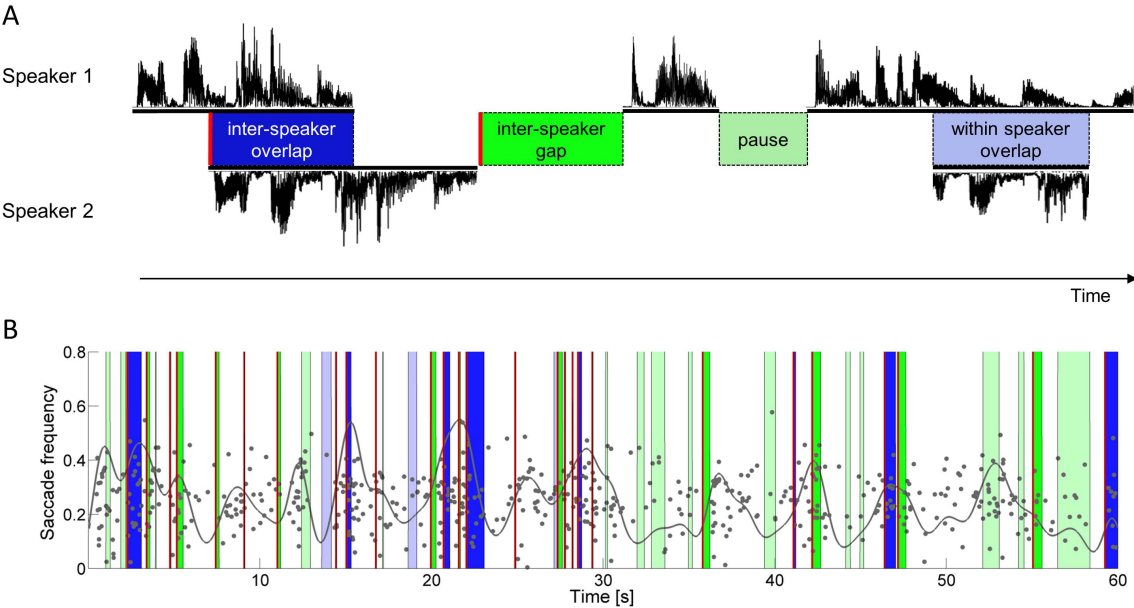


Figure 2. Frequency of gaze shifts

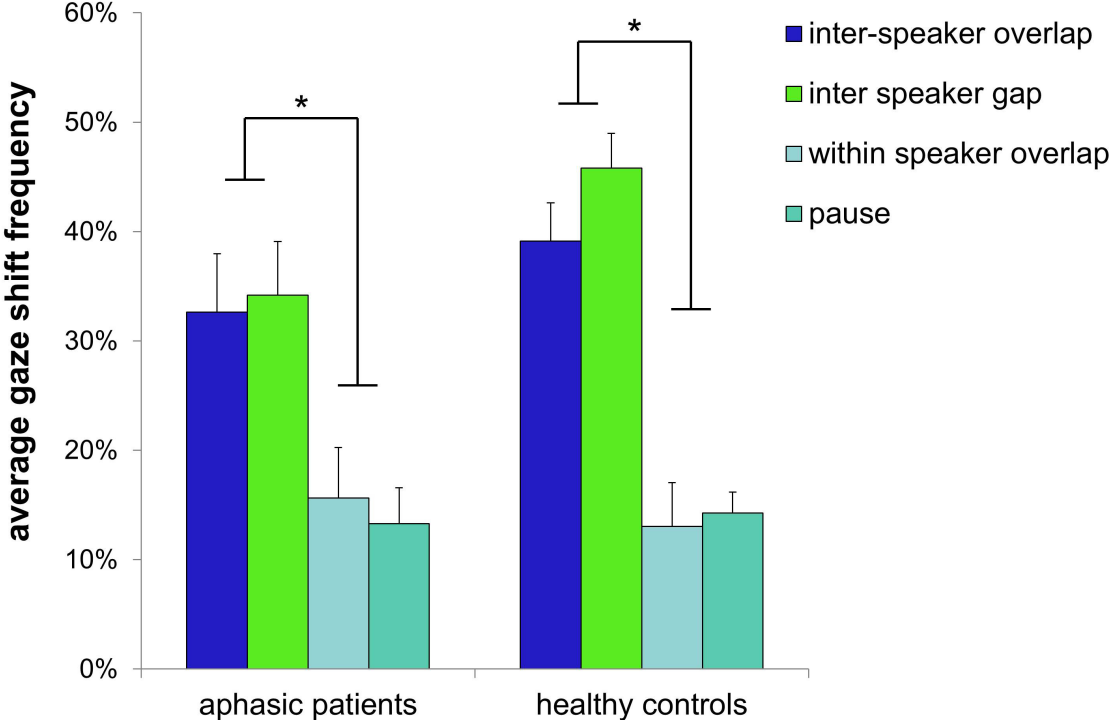


Figure 3. Gaze shift probability at turn transition

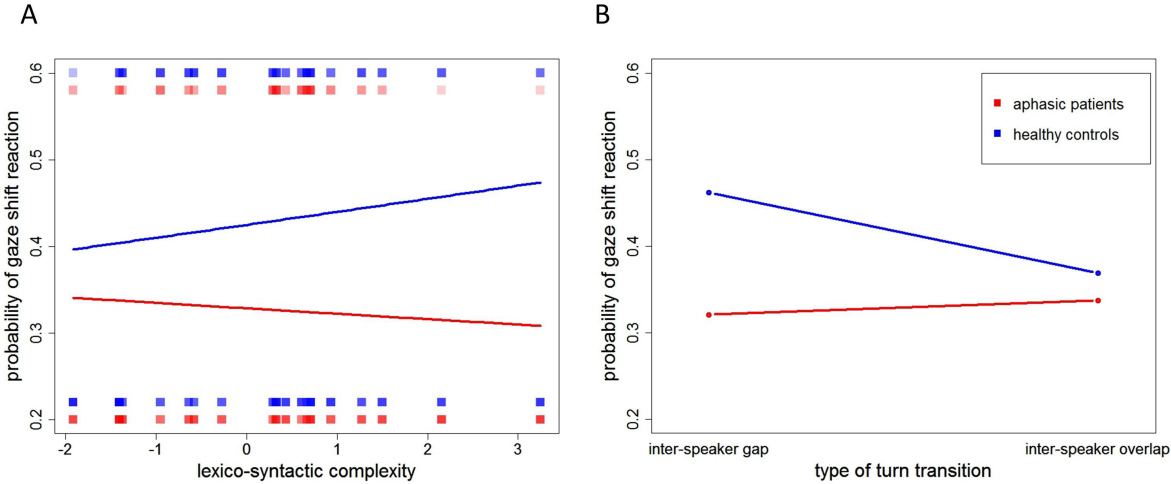


Figure 4. Timing of gaze shifts at turn transition.

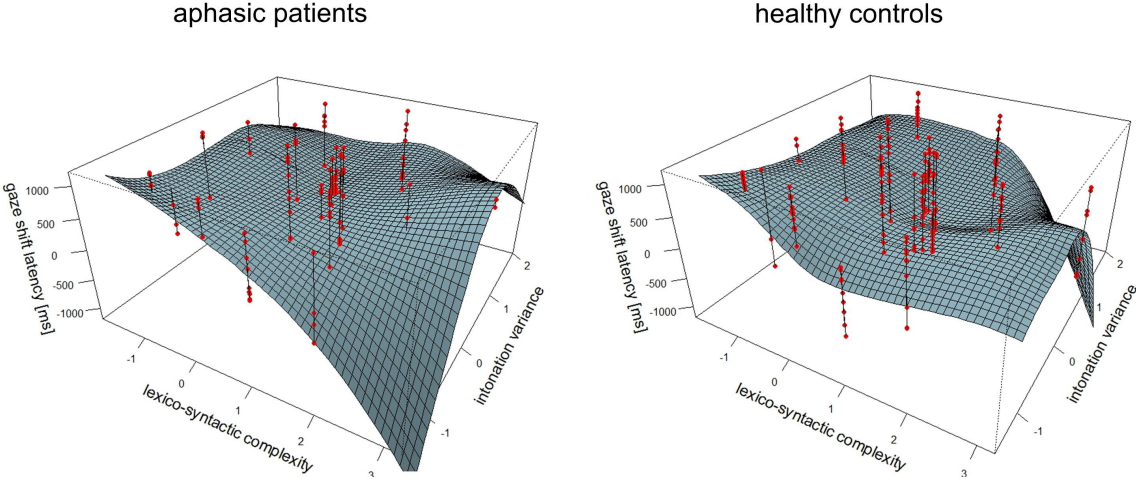


Figure 5. Voxel-based lesion-symptom mapping

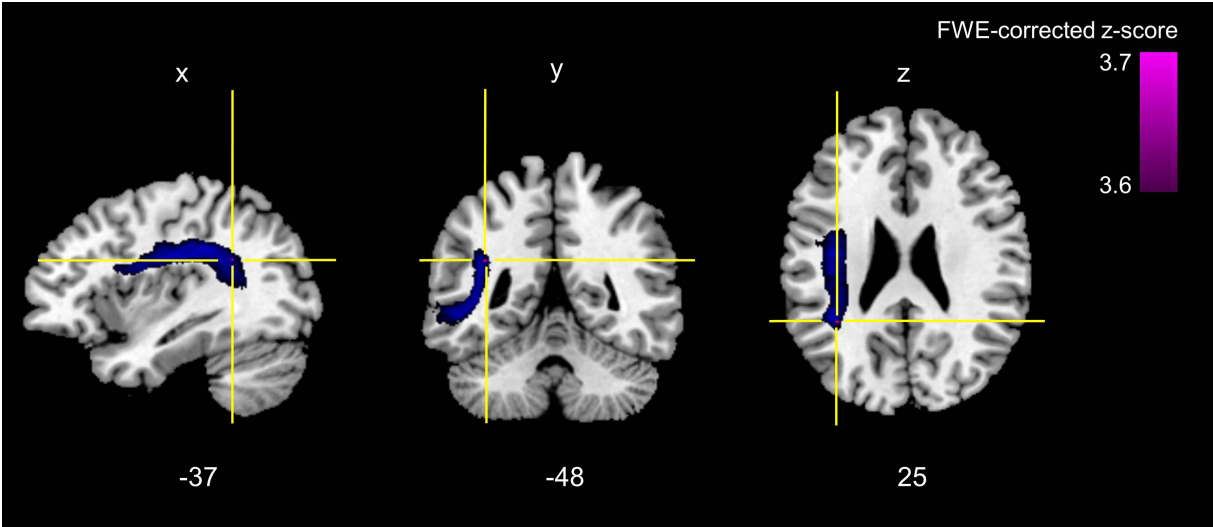


Figure 6. Lesion subtraction analysis

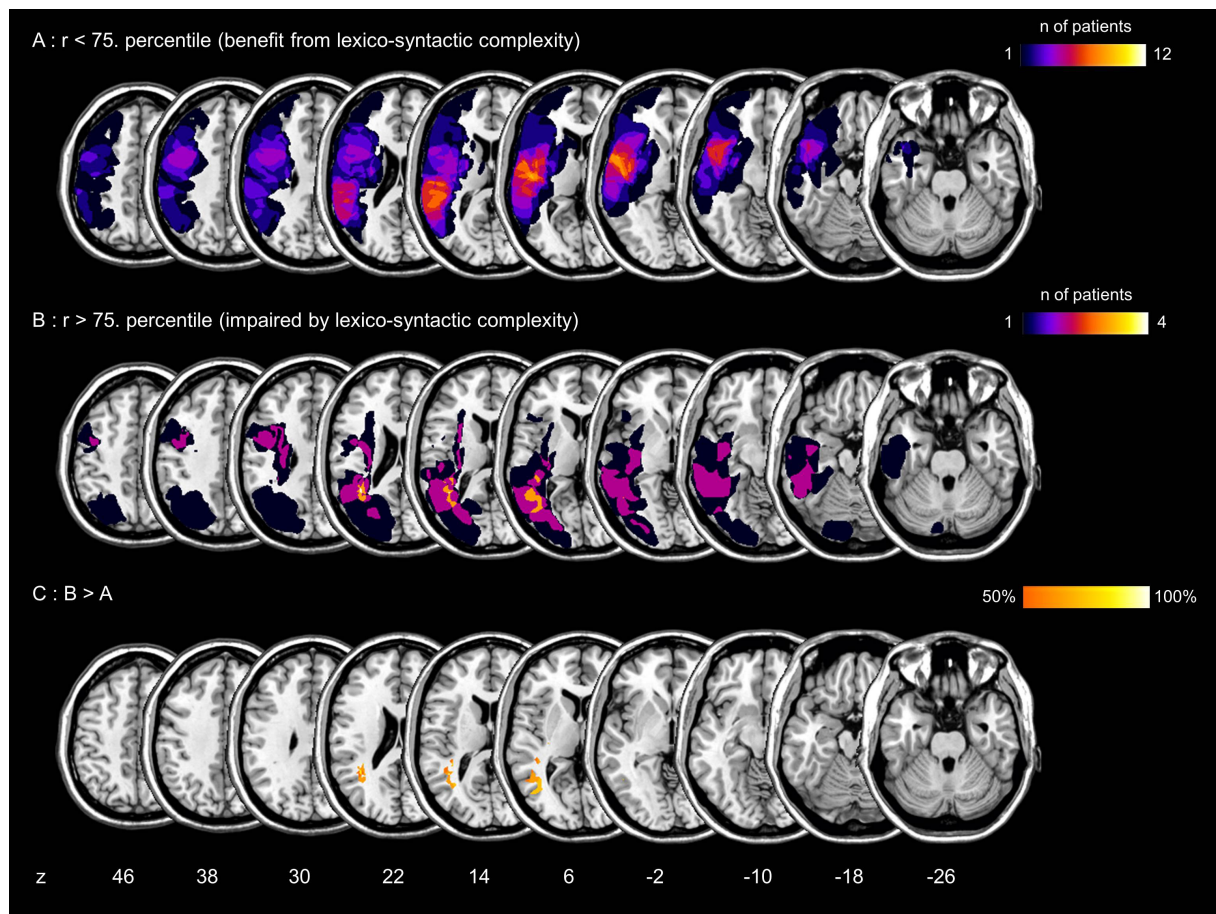


Table 1: Overview of video details

	video duration [s]	inter-speaker overlap			inter-speaker gap			pause			within-speaker overlap		
		number	\emptyset duration	<i>SD</i>	number	\emptyset duration	<i>SD</i>	number	\emptyset duration	<i>SD</i>	number	\emptyset duration	<i>SD</i>
video 1	132	14	390	381	36	395	360	27	634	526	6	634	509
video 2	128	17	371	393	16	291	270	21	460	301	11	461	205
video 3	121	16	335	347	10	191	130	16	320	172	11	551	334
video 4	124	3	265	238	15	240	171	44	384	227	2	415	209
mean	126	13	359	360	19	317	296	27	452	349	8	525	323

Note. \emptyset Duration = mean duration in ms; *SD* = Standard Deviation