Seçkin Arslan | Eren Gür | Claudia Felser

Predicting the sources of impaired wh-question comprehension in non-fluent aphasia

A crosslinguistic machine learning study on Turkish and German

Suggested citation referring to the original publication:
Cognitive Neuropsychology 34 (2017) 5, pp. 312–331
DOI http://dx.doi.org/10.1080/02643294.2017.1394284
ISSN (print) 0264-3294
ISSN (online) 1464-0627

Postprint archived at the Institutional Repository of the Potsdam University in:
Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 464
ISSN 1866-8364
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-412717
Predicting the sources of impaired *wh*-question comprehension in non-fluent aphasia: A cross-linguistic machine learning study on Turkish and German

Seçkin Arslana, Eren Gürb and Claudia Felsera

aPotsdam Research Institute for Multilingualism, University of Potsdam, Potsdam, Germany; bDepartment of Neurology, Hamidiye Şişli Etfal Research and Training Hospital, Şişli, Istanbul, Turkey

ABSTRACT

This study investigates the comprehension of *wh*-questions in individuals with aphasia (IWA) speaking Turkish, a non-*wh*-movement language, and German, a *wh*-movement language. We examined six German-speaking and 11 Turkish-speaking IWA using picture-pointing tasks. Findings from our experiments show that the Turkish IWA responded more accurately to both object *who* and object *which* questions than to subject questions, while the German IWA performed better for subject *which* questions than in all other conditions. Using random forest models, a machine learning technique used in tree-structured classification, on the individual data revealed that both the Turkish and German IWA’s response accuracy is largely predicted by the presence of overt and unambiguous case marking. We discuss our results with regard to different theoretical approaches to the comprehension of *wh*-questions in aphasia.

ARTICLE HISTORY

Received 26 January 2017
Revised 28 July 2017
Accepted 13 October 2017

KEYWORDS
Non-fluent aphasia; random forest algorithm; sentence comprehension; *wh*-in-situ; *wh*-questions; *wh*-movement

1. Introduction

Individuals with non-fluent aphasia (hereafter, IWA) often display difficulties comprehending sentences that involve changes to the canonical word order or require the computation of syntactic dependencies (e.g., Caramazza & Zurif, 1976). In particular, previous research has shown that the comprehension of *wh*-questions is challenging for IWA (e.g., Avrutin, 2000, 2006; Cho-Reyes & Thompson, 2012; Grodzinsky, 2000; Hanne, Burchert, & Vasishth, 2016; Hickok & Avrutin, 1996; Neuhaus & Penke, 2008; Stavrakaki & Kouvava, 2003; Thompson, Tait, Ballard, & Fix, 1999).

However, not all types of *wh*-question are equally impaired in aphasia, and no consensus has yet been reached about what might cause such selective deficits. Whilst some accounts hold that the presence of overt *wh*-movement makes certain *wh*-questions difficult, others point to difficulty integrating discourse-level information as a possible source of impaired comprehension. To test these claims, we conducted a cross-linguistic sentence-to-picture matching experiment on two groups IWA speaking Turkish, a non-*wh*-movement language, or German, a *wh*-movement language.

1.1. Linguistic background

The base word order of German is generally assumed to be subject–object–verb (SOV; e.g., Bach, 1962; den Besten, 1983), but due to the verb-second requirement the finite verb must raise to the complementizer position in German main clauses (e.g., Vikner, 1995), and as a result it will precede its object(s) in stylistically neutral declarative clauses. Assuming that subjects are base-generated verb phrase (VP) internally (Koopman & Sportiche, 1991), the derivation of SVO order, as, for example, in subject *wh*-questions such as (1), involves not only verb raising but also subject movement to (Specifier (Spec), Complementizer Phrase (CP)), as indicated (in simplified form) in (1)2 (NOM = nominative; ACC = accusative; C = complementizer; IP = inflection phrase; t = trace; subj = subject).

(1) Wer küsst den Mann? 
who[NOM] kisses the[ACC] man 
Who is kissing the man? 
(1′) [CP Wer[subj] [C ‘[CP Küsst, t][IP t[subj] . . . [VP t[subj] [V[ACC] den Mann] [V[ACC] [t, t] ] ]]]]]

German being a case-marking language, in (1) the direct object noun phrase (NP) *den Mann* carries

CONTACT Seçkin Arslan seckin.arslan@uni-potsdam.de; seckin1984@gmail.com

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.
accusative case whereas the subject wh-pronoun wer carries nominative case.

Note that even though multiple derivational steps may be involved in generating German subject questions, the original relative ordering of subject and object (S > O) is preserved. This is not the case for German object questions such as (2), which require overt movement of the object wh-phrase across the subject to (Spec, CP), as indicated in (2').

\[
\begin{align*}
(2) & \quad \text{Wen küsst der Mann?} \\
& \quad \text{who}_{\text{ACC}} \quad \text{kiss}\text{}_\text{subject} \quad \text{the}\text{ }\text{man}\text{ }\text{NOM}
\end{align*}
\]

In (2), the wh-pronoun wen carries accusative case, which provides an unambiguous cue that it functions grammatically as an object, whereas the subject NP der Mann is unambiguously marked as nominative. Note that due to case syncretism for feminine NPs, both nominative and accusative feminine definite NPs are introduced by the article die.

Turkish employs free word order to a greater extent than German does. The canonical word order is taken to be SOV, yet other variations are allowed (Taylan, 1984). Unlike German, Turkish allows for wh-in-situ, where wh-phrases remain in their base positions (Akar, 1990; Özyoğ, 2009). See Examples 3 and 4 for illustration (PRES.PROG. = present progressive).

\[
\begin{align*}
(3) & \quad \text{Kim adamı öpüyor?} \\
& \quad \text{who}_{\text{ACC}} \quad \text{man}_{\text{ACC}} \quad \text{kiss}\text{}_\text{PRES.PROG.}
\end{align*}
\]

Example 3 shows a subject who question. The subject appears in sentence-initial position by default in Turkish. Like their German counterparts, Turkish subject wh-questions thus do not require any word-order-changing wh-movement. Example 4 shows an object who question. A preferred location for theme objects in Turkish is the immediate pre-verbal position. However, when the object is being questioned, unlike in German no wh-movement will apply. Instead, a wh-phrase appears in the canonical object position. Notice that overt accusative case marking is used on object wh-phrases depending on the verb semantics.

1.2. Studies on the comprehension of wh-questions in aphasia

According to a number of studies, IWA’s difficulties in comprehending certain types of sentences, including wh-questions, reflect problems building target-like syntactic representations. In German and English, object questions such as Which movie did you see last night? require fronting of the object NP. In generative-transformational theory (Chomsky, 1981) traces are assumed to be left behind in an NP’s base-generated or canonical position as well as in all intermediate positions that the NP may have moved through. Grodzinsky’s (2000) trace deletion hypothesis (TDH) holds that traces are absent from the syntactic representations of IWA. If we adopt the VP-internal subject hypothesis (Koopman & Sportiche, 1991), then deriving object questions in wh-movement languages requires both the object and the subject NP to undergo movement out of the VP; see Example 2’ above for illustration. If, as Grodzinsky claims, both subject and object traces are missing from IWA’s syntactic representations, then neither NP will be able to receive a thematic role from the verb. Comprehenders may then have to resort to simply guessing the semantic role of each NP. According to Grodzinsky (1995), IWA may use a default strategy of assigning the agent role to the sentence or clause-initial NP. This strategy will yield the correct semantic role assignment in subject questions but result in the misinterpretation of object questions, where initial wh-expressions represent themes that query the object. In short, from the perspective of the TDH, IWA are expected to perform relatively better interpreting subject than object questions.

According to another view, IWA’s difficulties with object questions are due to the fact that anther NP intervenes between the wh-word and its trace (Garraffa & Grillo, 2008; Sheppard, Welsh, Love, & Shapiro, 2015). Extending the intervention hypothesis to other kinds of A-bar movement, Friedmann, Rizzi, and Belletti (2017) report that a group of Hebrew-speaking IWA failed to comprehend object-topicalized sentences in comparison to non-object-first sentences. This is consistent with the authors’ hypothesis that in aphasia, lexically restricted NPs intervening between a fronted constituent and its trace can act as interveners. The authors found that overt case marking on object NPs did not help IWA to better
interpret sentences with fronted NPs, leading them to hypothesize that case features are not in fact involved in the computation of A-bar dependencies. This is also what Burchert, De Bleser, and Sonntag (2003) showed for some of their German IWA who had difficulty comprehending non-canonical sentences with fronted NPs even where case marking on objects provided unambiguous cues to thematic role assignment. The authors found no overall difference between sentences with ambiguous and unambiguous case marking, although for the latter two of their IWA seemed to show better performance. Both the intervention hypothesis and the TDH point towards the same underlying problem in IWA’s comprehension difficulties: the presence of overt—and, in particular, of argument-order-changing—wh-movement.

One may, however, wonder whether the presence of overt wh-movement actually matters, as not all languages necessarily employ this feature. In so-called wh-in situ languages, object wh-phrases may outscope subject phrases at the semantic level despite appearing in a hierarchically lower surface position. The above question was addressed by van der Meulen, Bastiaanse, and Rooryck (2005), who assessed French IWA, a language that has both wh-in-situ and wh-movement object questions, using a picture-pointing task. French IWA performed more poorly on wh-movement object questions than on their wh-in-situ counterparts. Thus, the authors identified overt wh-movement as a critical factor leading to compromised comprehension of wh-questions in aphasia. Furthermore, Drai and Grodzinsky (2006) conducted a meta-analysis with 69 IWA from different studies on a range of languages including Dutch, English, Hebrew, German, Japanese, Korean, and Spanish, examining comprehension of relative clauses, active/passive affirmative sentences, and wh-questions. Besides other factors, the authors examined effects of movement (including movement assumed to apply in affirmative SVO sentences in German and Dutch) on IWA’s sentence comprehension. The authors identified the presence of non-string-vacuous movement as a significant factor contributing to IWA’s sentence comprehension problems.

In summary, from the point of view of representational deficits accounts, IWA perform poorly in questions requiring wh-movement because they fail to associate a fronted NP with its trace, due to either the presence of an intervening NP or a more general inability to compute traces. This assumption, however, has not been supported by recent studies using time-course-sensitive measurements. Using the visual-world paradigm to monitor participants’ eye-movements, Dickey, Choy, and Thompson (2007) and Thompson and Choy (2009) for English wh-questions and object clefts, and Hanne et al. (2016) for German object and subject who questions showed that IWA’s eye movements patterned with those of their controls for correctly answered trials. The data indicated that IWA were virtually perfectly able to turn their gaze to a target picture (i.e., the picture showing the theme argument) during the processing of the verb and post-verbal regions of object questions, despite showing delayed and less accurate end-of-trial responses. These findings have been taken to suggest that IWA are relatively spared in computing wh-dependencies (e.g., Dickey et al., 2007), with non-fluent IWA showing sensitivity to fronted elements’ base positions (see also Blumstein et al., 1998). According to such processing deficits accounts, IWA’s syntactic representations are not impaired per se, but their processing of certain syntactic structures is delayed or weakened.

Several studies have shown that IWA experience comprehension difficulty when sentence interpretation requires the integration of discourse-level information, which is assumed to be computationally costly in aphasia (e.g., Avrutin, 2000, 2006; Bos, Dragoy, Avrutin, Iskra, & Bastiaanse, 2014; Fyndanis, Varlokosta, & Tsapkini, 2010; Hickok & Avrutin, 1996; Nenarctzni, Varlokosta, Papadopoulou, & Bastiaanse, 2014; Salis & Edwards, 2008). There is cross-linguistic evidence showing that both subject and object which questions are more susceptible to aphasic impairments than who questions; see, for instance, Nenarctzni et al. (2014) for Greek, Avrutin (2000, 2006), Hickok and Avrutin (1996), and Salis and Edwards (2008) for English, and Bos et al. (2014) for Russian. Within which questions, object questions tend to be comprehended less well than subject ones. In order to account for the observed asymmetry between which- and who-type questions, Avrutin (2000, 2006) suggests that, based on the theoretical framework of Pesetsky (1987), which phrases require linking to specific sets of discourse entities, whilst wh-pronouns such as who are non-referential and non-discourse-linked. Therefore, computing mental representations for which questions requires more
computational resources than computing who questions. These additional resource requirements exceed IWA’s computational capacity, leading to problems processing which questions.

Furthermore, Halliwell (2004) studied IWA’s sentence comprehension in Korean, a wh-in-situ language, and found that IWA performed above chance both for structures that require syntactic movement (e.g., passives) and for those that do not (e.g., subject and object questions). Kljajevic and Murasugi (2010) investigated the comprehension of wh-questions in a group of IWA speaking Croatian, which allows for greater word-order variation than the Indo-European languages explored previously. A subset of their participants attained better scores in comprehending object questions than subject ones, while the others showed relatively spared comprehension across all object and subject wh-conditions. The authors reported that morphologically marked accusative case on the wh-phrase constitutes an important cue that Croatian IWA rely on to identify the agent and theme of a sentence (contra what has been reported by Friedmann et al., 2017, for Hebrew-speaking IWA).

Summarizing, most previous research has taken linguistic factors such as the presence of overt wh-movement (Drai & Grodzinsky, 2006), or both movement and referentiality of wh-phrases (Avrutin, 2000, 2006; Bos et al., 2014; Garraffa & Grillo, 2008; Grodzinsky, 1995; Hickok & Avrutin, 1996) as critical factors in accounting for aphasic comprehension difficulties. However, virtually all the above-mentioned studies also report large differences between individual patients. For instance, Thompson et al. (1999) showed that each one of four IWA demonstrated different directions of impairments in comprehending wh-questions in English. It is still far from clear which individual factors, besides linguistic constraints, contribute to IWA’s difficulties in comprehending wh-questions.

While some IWA had less difficulty comprehending both which and who subject questions than object questions, others performed virtually perfectly in all conditions except for object which questions. Still others performed below chance level in all conditions. Neuhaus and Penke (2008) proposed the following scaling of difficulty for German IWA: Subject who questions are retained best, followed by subject which and object who questions. Object which questions are the most severely affected type.

Hanne, Burchert, De Bleser, and Vasishth (2015), using a visual-world paradigm, examined the role of unambiguous case marking in a group of German IWA’s sentence comprehension abilities in non-canonical sentences. The authors report that their IWA showed reduced sensitivity to unambiguous case marking in their offline responses to non-canonically ordered sentences, replicating Burchert et al. (2003) findings. However, Hanne et al.’s (2015) online data showed that their IWA were able to process unambiguous case cues successfully, indicating that German IWA are sensitive to case marking as an interpretation cue during processing.

Hanne et al. (2016) studied a group of German IWA’s comprehension of object and subject who questions using a visual-world paradigm. Their data revealed that a subset of their participants performed poorly in both conditions while some others performed better either for subject who questions or for object ones. The authors reported an advantage in IWA’s response times, however, for object questions compared to subject ones. As their participants’ eye movements patterned with those of unimpaired controls, Hanne et al. (2016) argued that IWA retain the ability to process wh-structures online despite showing decreased offline response accuracy.

Nothing as yet has been explored about the comprehension of wh-questions in Turkish aphasia. However, results from the few existing studies on Turkish indicate that the processing of certain syntactic structures, such as object and subject relative clauses, is impaired (Arslan, Bamyaci, & Bastiaanse, 2016; Aydin, 2007; MacWhinney, Osmán-Sági, & Slobin, 1991; Yarbay-Duman, Altınoz, & Bastiaanse, 2011). For instance, MacWhinney et al. (1991) studied Turkish IWA’s interpretation of several kinds of sentences in which word order and case marking on the NPs were manipulated. Their data indicated that Turkish IWA had more difficulty

1.3. Previous studies on German and Turkish IWA

Neuhaus and Penke (2008) studied German-speaking IWA’s comprehension of wh-questions using a picture-pointing task. Their data showed that object which questions were more prone to comprehension deficits in non-fluent aphasia than subject which questions, and that object who questions were more affected than their subject counterparts. However, important individual differences were observed. For instance, Thompson et al. (1999) showed that each one of four IWA demonstrated different directions of impairments in comprehending wh-questions in English. It is still far from clear which individual factors, besides linguistic constraints, contribute to IWA’s difficulties in comprehending wh-questions.
comprehending object-topicalized sentences in which the clause-initial object NP carried accusative case than sentences with object NPs positioned in their canonical order, suggesting that Turkish aphasia might involve a reduced ability to use morphological cues during sentence comprehension. Yarbay-Duman et al. (2011), using a sentence-to-picture matching task, studied Turkish IWA’s comprehension of declarative base order and scrambled sentences, subject relatives, object relatives, and passives. Turkish employs an assortment of case marking on agents and themes in those structures to signal thematic role assignments. For instance, base order or scrambled declaratives and subject relatives all have nominative agents and accusative-marked themes, whilst passives constructed with by-phrases have nominative on both the agent and theme. The data showed that their IWA found base-order declaratives easier to comprehend than object scrambling and subject relatives, with object relatives and passives being the most problematic sentence types. The authors conclude that structures that follow base order (i.e., SOV sentences with nominative agents and accusative-marked themes) are less vulnerable in aphasia than structures in which case assignment is less transparent, as, for example, in passives.

In summary, it has been shown that (a) object questions tend to be more affected in comprehension than subject questions in IWA speaking wh-movement languages, (b) the presence of unambiguous accusative case marking seems to help Croatian- and Turkish-speaking IWA to some degree but does not improve German- or Hebrew-speaking IWA’s sentence comprehension ability.

In the current study, we conducted a sentence comprehension experiment using a picture-pointing task administered to groups of Turkish and German IWA and to control groups of non-brain-damaged (NBD) Turkish and German monolingual individuals. The following research questions are addressed:

- Do German- and Turkish-speaking IWA differ in their ability to comprehend wh-questions?
- Which linguistic and individual factors significantly predict impairments in Turkish and German IWA’s comprehension of wh-questions?

The above-mentioned theoretical accounts predict German and Turkish IWA to perform differently in comprehending wh-questions. In particular, the representational deficit accounts we have reviewed predict Turkish IWA to be spared in comprehending wh-questions generally, whereas German IWA are expected to be challenged by the presence of wh-movement, especially in object questions as these involve movement that changes the canonical argument ordering (e.g., Drai & Grodzinsky, 2006). The intervention hypothesis (e.g., Friedmann et al., 2017) furthermore predicts that the presence of unambiguous case marking will not help IWA to interpret wh-movement structures. The discourse-linking hypothesis (e.g., Avrutin (2000, 2006), on the other hand, would predict both German and Turkish IWA to perform alike in responding more poorly to which questions than to who questions. This is based on the assumption that discourse-based processes should apply similarly in both the languages irrespectively of wh-movement.

2. Method

2.1. Participants

Our participants included six German-speaking and 11 Turkish-speaking individuals who suffer from aphasia. Table 1 provides an overview of our participants’ demographic and diagnostic details, and Table 2 shows the IWA’s scores from their most recent language assessment profile for different aphasia tests.

The Turkish IWA were recruited at the Anadolu University Research and Rehabilitation Centre for Language and Speech Pathology (DILKOM) in Eskisehir, and at the Şişli Hamidiye Etfal Research and Training Hospital in Istanbul. Their aphasia diagnoses were confirmed using the Aphasia Language Assessment Test (Maviş & Toğram, 2009). The German IWA were recruited through the patient database of the University of Potsdam and tested either at their homes or at an affiliated speech and language therapy clinic. The Aachen Aphasia Test was used for the aphasia diagnoses for German (Huber, Poeck, Weniger, & Willmes, 1983). In order for us to obtain a directly comparable score on auditory comprehension, both groups of participants were examined using the Token Test (De Renzi & Faglioni, 1978). For the German IWA the Token Test scores were taken from their most recent Aachen Aphasia Test (AAT) assessment (all < 7 months), and for the Turkish IWA the Token Test was administered prior to the experimental sessions.
In addition, two groups of German (n = 17, M<sub>age</sub> = 61.82 years, range = 52–68 years) and Turkish (n = 13, M<sub>age</sub> = 62.38 years, range = 55–67 years) non-brain-damaged native speakers (NBDs) took part in the experiment as control groups. The NBDs were of comparable age to the IWA in both the Turkish [Welch t (1669.403) = 1.5327, p = .125] and German groups [Welch t(658.152) = 1.3495, p = .177]. The NBDs had no psychiatric or neurological disorders. All participants signed a consent form and were paid 8 Euros per hour for their participation.

### 2.2. Materials

Our sentence stimuli contained 24 interrogative sentences in four conditions: ObjectWhich, ObjectWho,
To construct these stimuli, 12 verbs were chosen (see Appendix A) and were used in interrogative sentences (see Appendix B) with reversible agents and themes (e.g., *Which man kissed the woman*?). In half of the items in each condition (n = 12), the agent was male and the theme was female, and in the other half, the agent–theme pairs were reversed (e.g., *Which woman kissed the man*?). In the German experiment, two conditions (Object-Which and ObjectWho) involved wh-fronting of the object, while in the other two conditions (SubjectWhich and SubjectWho), the canonical ordering of arguments was preserved. In the Turkish experiment, by contrast, all experimental sentence stimuli contained in-situ wh-phrases and showed SVO word order.

To display the actions described by the four wh-question types visually, 96 photos were used as visual stimuli. In each of these photos three human referents were shown performing an action, as illustrated in Figure 1. For instance, the action "to push" was photographed in four different versions. In two versions, a male actor appeared in the middle, and two female actors to the left and right sides of him performing the action directed either to the left or to the right-hand side. These four versions of photos were counterbalanced throughout the experimental items in order for us to control for potential effects of action direction and/or gender bias.

In the ObjectWhich and SubjectWhich conditions, the sentence stimuli contained the wh-phrases *which man* or *which woman* depending on the gender of the theme/agent being questioned, requiring the presence of at least two male or female actors in the corresponding visual stimuli. Although in the Object-Who and SubjectWho conditions, the sentence stimuli (e.g., *Who pulls the man*) are visually depictable with a female and a male actor only, we presented these items with photos where there were three actors, similar to the ObjectWhich and SubjectWhich conditions. This was to be able to maintain the internal consistency of our experimental materials and to avoid participants’ developing strategies based on the visual stimuli.

### 2.3 Procedure

Each participant was tested individually either at their home or in a dedicated room at the speech-therapy clinic or hospital. The visual stimuli described above were shown to the participants in random order with one photo on each page with the sentence-stimuli printed underneath. The experimenter read aloud each sentence to the participant and asked them to answer
the question by pointing to the person in the photo that corresponded to the agent (for subject questions) or theme (for object questions) of the event described. The experiment was not timed. Three practice items were provided prior to the experiment and were repeated if necessary until the task was fully understood. The overall experiment took around 60 min for the IWA to complete and around 20 min for NBDs. Aphasia assessment tests (except for the Token Test for Turkish IWA) were administered in separate sessions.

The experiments reported here were carried out in accordance with The Code of Ethics of the World Medical Association and the Declaration of Helsinki, and were approved by the ethics committee of the University of Potsdam (application number: 54/2015).

2.4. Scoring and analyses
An answer was scored as accurate when the participant pointed to the correct referent of the wh-phrase or pronoun. Answers were scored as inaccurate when the participant pointed to an incorrect referent.

2.4.1. Overall group analyses
The overall group performances were analysed using mixed-effects logistic regression using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015); post hoc comparisons were computed with the Tukey’s test using the “multcomp” package. The statistical significance threshold was set to $p < .05$ (or $t > \pm 2.00$ where applicable). In the regression analyses participants and items were added as random intercepts and slopes following Baayen, Davidson, and Bates (2008). A prototype model was first built, and then each factor and/or intercepts and slopes were successively removed, and the best fitting model was reported based on the Akaike information criterion (AIC).

2.4.2. Determining predictors of sentence comprehension failures through the random forest machine learning algorithm
We further analysed our data using the random forest (RF) algorithm, a machine learning technique for non-parametric tree-structured classification (Breiman, 2001). A RF is an ensemble of decision trees, each of which is built with the recursive partitioning principle used in multivariate data analyses (Strobl, Malley, & Tutz, 2009). One of the most important reasons we used RF models to further analyse our data is that logistic mixed-effect regression models are vulnerable to correlations between variables. In RF models,
however, correlations are not an issue, and they can handle over 50 different variables as predictors, as well as interactions between these predictors, at once without compromising model accuracy (see Tagliamonte & Baayen, 2012, for discussions).

RFs were implemented with the following steps (see Appendix C for the example R codes that are used for each of those steps):

1. **Variable selection:** We implemented RFs on Turkish and German IWA’s accuracy of responses to our experimental tasks. Two sorts of variables were used: The dependent variable was participants’ response accuracy (structured as binomial data signalling whether or not a participant responded accurately), and predicting variables were potential determiners of IWA’s accurate responses. The following predicting variables were selected:
   i. **Demographic predictors**
   - Age and years of education of our IWA
   ii. **Linguistic and experimental predictors**
   - Case of *wh*-elements (nominative vs. accusative): In both Turkish and German, *wh*-elements in object questions carry accusative case while *wh*-elements in subject questions carry nominative case.
   - Case of non-*wh* NPs (nominative vs. accusative): In object questions, the NP associated with the agent role carries nominative case, while in subject questions the NP associated with the theme role carries accusative case.
   - Case ambiguity—that is, whether or not the sentence stimulus contains an ambiguous case cue. This only applies to German sentences. Half of the ObjectWho and SubjectWho items were marked as unambiguous as these sentences contained masculine non-*wh* NPs, which are unambiguously case-marked. The remaining items were classified as ambiguous due to the presence of a feminine NP or *wh*-phrase, which are formally ambiguous between nominative and accusative.
   - Case syncretism of the *wh*-phrases—that is, whether or not case distinctions on *wh*-phrases are realized by different forms. This again only applies in German.
   - *Wh*-movement—that is, whether or not a stimulus sentence displays argument-order-changing *wh*-movement.
   - Gender of the theme argument—that is, whether the person referent who underwent the action in the experimental sentence stimuli was male or female.
   iii. **Aphasiological predictors**
   - Post onset: number of months following the onset of aphasia.
   - Token Test scores: We used the results from the Token Test as a predictor because this test is used for examining the comprehension of commands in which complex grammatical structures are only minimally involved (at least for German and Turkish). The Token Test scores may thus provide us with a unified index of how severely our IWA’s comprehension ability is affected without being confounded by typological differences between the Turkish and German grammars.
   - Percentage word naming, word repetition, and sentence and word comprehension scores.

2. **Random forest implementation:** We implemented RF using the “cforest” function of the “party” package in R (Hothorn, Hornik, & Zeileis, 2006), enabling the algorithm to learn from the data by assembling unbiased decision trees. A RF implementation was considered to be accurate when the C ratio was equal to or greater than 0.80 (see Tagliamonte & Baayen, 2012).

3. **Variable importance:** The importance of predicting variables was computed via the “varImp” function following Strobl, Hothorn, and Zeileis (2009). In our variable importance calculations, conditional permutations were preferred because the predicting variables strongly correlated with each other. Predicting variables below or close to a “0” z-value of importance were considered as uninformative and were dropped from further analyses.

4. **Conditional inference trees:** The conditional inference trees reported in this study were plotted using the ‘ctree’ function of the “party” package
in R (Strobl, Hothorn, et al., 2009). Following previous studies that employed RF on aphasia data (e.g., de Aguiar, Bastiaanse, & Miceli, 2016), we built the conditional inference trees first with all predictors, and then each of the uninformative variables was removed from the model if their removal did not reduce the RF classification accuracy. The most pertinent conditional inference tree was then reported.

3. Results

3.1. Overall results

Table 4 shows the individual and group accuracies of our participants’ responses. Two logistic mixed-effects regression models were computed for each language group to compare IWA to NBD. The outputs from these models show that the IWA performed more poorly than the NBDs in both the Turkish and the German groups (Turkish: $\beta = 4.588$, $SE = 0.470$, $z = 9.750$, $p < .001$; German: $\beta = 4.3689$, $SE = 0.361$, $z = 12.103$, $p < .001$). The Turkish NBD controls performed with $98.8\%$ ($SD = 10.5$) overall response accuracy, and the German NBDs performed at $98.0\%$ ($SD = 13.8$) accurate in responding to the task. Therefore, the data from the NBDs are not further analysed.

Table 5 provides the outputs from the main mixed-effects logistic regression model computed with the data from our Turkish and German IWA. In this model, condition and language group were added as fixed factors, and post onset and severity of the IWA were added as factors in order for us to assess whether or not effects observed in their responses were modulated by individual aphasia-related factors. Additionally, the model accommodated participants and items as random intercepts. The outputs show significant effects of condition (with the ObjectWho condition only approaching significance), of language group, and of the interactions between condition and language group (Table 5). The model also showed significant effects of aphasia severity for the IWA but not of post-onset time. Since there were significant interactions between language group and condition, response accuracies from our German and Turkish IWA were analysed separately using post hoc Tukey tests.

The Turkish IWA performed better in the Object-Who condition than in the ObjectWho ($\beta = -0.637$, $SE = 0.201$, $z = -3.16$, $p = .008$), SubjectWhich ($\beta = -1.196$, $SE = 0.200$, $z = -5.95$, $p < .001$), and Subject-Who conditions ($\beta = -1.142$, $SE = 0.200$, $z = -5.96$, $p < .001$). They performed more accurately in the ObjectWho condition than in both the SubjectWhich ($\beta = -0.559$, $SE = 0.190$, $z = -2.93$, $p = .017$) and SubjectWho conditions ($\beta = -0.505$, $SE = 0.190$, $z = -2.652$, $p = .04$). However, the Turkish IWA’s response accuracy did not differ between the SubjectWho and SubjectWhich conditions ($\beta = 0.053$, $SE = 0.188$, $Z = 0.284$, $p = .99$).

Table 4. Accuracy proportions of the Turkish and German IWA’s comprehension of wh-questions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Object Which</th>
<th>Object Who</th>
<th>Subject Which</th>
<th>Subject Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>German IWA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE-A01</td>
<td>.33 (.48)</td>
<td>.38 (.49)</td>
<td>.88 (.33)</td>
<td>.87 (.34)</td>
</tr>
<tr>
<td>DE-A02</td>
<td>.17 (.38)</td>
<td>.50 (.51)</td>
<td>.17 (.38)</td>
<td>.28 (.45)</td>
</tr>
<tr>
<td>DE-A03</td>
<td>.29 (.46)</td>
<td>.42 (.50)</td>
<td>.67 (.48)</td>
<td>.17 (.38)</td>
</tr>
<tr>
<td>DE-A04</td>
<td>.38 (.49)</td>
<td>.33 (.48)</td>
<td>.58 (.50)</td>
<td>.38 (.49)</td>
</tr>
<tr>
<td>DE-A05</td>
<td>.50 (.51)</td>
<td>.33 (.48)</td>
<td>.75 (.44)</td>
<td>.58 (.51)</td>
</tr>
<tr>
<td>DE-A06</td>
<td>.71 (.46)</td>
<td>.46 (.51)</td>
<td>.67 (.48)</td>
<td>.21 (.41)</td>
</tr>
<tr>
<td>Mean</td>
<td>.40 (.49)</td>
<td>.40 (.41)</td>
<td>.62 (.48)</td>
<td>.41 (.49)</td>
</tr>
<tr>
<td>Turkish IWA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR-A01</td>
<td>.54 (.51)</td>
<td>.33 (.48)</td>
<td>.25 (.44)</td>
<td>.21 (.41)</td>
</tr>
<tr>
<td>TR-A02</td>
<td>.63 (.49)</td>
<td>.63 (.49)</td>
<td>.25 (.44)</td>
<td>.29 (.46)</td>
</tr>
<tr>
<td>TR-A03</td>
<td>.75 (.44)</td>
<td>.50 (.51)</td>
<td>.50 (.51)</td>
<td>.54 (.51)</td>
</tr>
<tr>
<td>TR-A04</td>
<td>.83 (.38)</td>
<td>.71 (.46)</td>
<td>.63 (.49)</td>
<td>.58 (.51)</td>
</tr>
<tr>
<td>TR-A05</td>
<td>.83 (.39)</td>
<td>.83 (.38)</td>
<td>.58 (.50)</td>
<td>.46 (.51)</td>
</tr>
<tr>
<td>TR-A06</td>
<td>.96 (.20)</td>
<td>.67 (.48)</td>
<td>.75 (.44)</td>
<td>.87 (.20)</td>
</tr>
<tr>
<td>TR-A07</td>
<td>.50 (.51)</td>
<td>.63 (.49)</td>
<td>.36 (.49)</td>
<td>.46 (.51)</td>
</tr>
<tr>
<td>TR-A08</td>
<td>.75 (.44)</td>
<td>.38 (.49)</td>
<td>.38 (.49)</td>
<td>.29 (.46)</td>
</tr>
<tr>
<td>TR-A09</td>
<td>.42 (.52)</td>
<td>.42 (.50)</td>
<td>.46 (.50)</td>
<td>.21 (.41)</td>
</tr>
<tr>
<td>TR-A10</td>
<td>.96 (.20)</td>
<td>.92 (.28)</td>
<td>.83 (.38)</td>
<td>.79 (.41)</td>
</tr>
<tr>
<td>TR-A11</td>
<td>.92 (.28)</td>
<td>.75 (.44)</td>
<td>.46 (.40)</td>
<td>.79 (.40)</td>
</tr>
<tr>
<td>Mean</td>
<td>.73 (.44)</td>
<td>.61 (.48)</td>
<td>.50 (.50)</td>
<td>.51 (.50)</td>
</tr>
</tbody>
</table>

Note: Standard deviations in parentheses. IWA = individuals with non-fluent aphasia; DE = German agrammatic speaker; TR = Turkish agrammatic speaker.

Table 5. Outputs from the main generalized mixed-effects regression model for the aphasia data.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.250</td>
<td>3.839</td>
<td>-0.582</td>
</tr>
<tr>
<td>Post onset</td>
<td>2.954e</td>
<td>1.430e</td>
<td>0.201</td>
</tr>
<tr>
<td>Severity</td>
<td>2.961</td>
<td>4.084e</td>
<td>7.251</td>
</tr>
<tr>
<td>Condition (ObjectWho)</td>
<td>6.993</td>
<td>3.909</td>
<td>1.789</td>
</tr>
<tr>
<td>Condition (SubjectWhich)</td>
<td>1.410</td>
<td>3.238</td>
<td>4.356</td>
</tr>
<tr>
<td>Language (Turkish)</td>
<td>6.682</td>
<td>3.013</td>
<td>2.218</td>
</tr>
<tr>
<td>Language (German)</td>
<td>1.792</td>
<td>2.999</td>
<td>5.976</td>
</tr>
<tr>
<td>Language (ObjectWho) × Condition (SubjectWhich)</td>
<td>-1.333</td>
<td>4.395</td>
<td>-0.303</td>
</tr>
<tr>
<td>Language (Condition (SubjectWho) × Language)</td>
<td>-2.598</td>
<td>3.810</td>
<td>-0.681</td>
</tr>
<tr>
<td>Language (Condition (SubjectWho) × Language)</td>
<td>-1.803</td>
<td>3.618</td>
<td>-0.494</td>
</tr>
</tbody>
</table>

Note: Code in R: glmer(Accuracy ~ Postonset + Severity + Condition + Language + (1|Item) + (1|Participant), data = accuracyGLM, family = binomial).
As more recent onset of aphasia may be differently impacting on comprehension processes, we repeated our analyses on the Turkish IWA who had a post-onset time of less than six months (TR-A06, TR-A09, TR-A10, and TR-A11) and on those with a post-onset time of more than six months (TR-A01, TR-A02, TR-A03, TR-A04, TR-A05, TR-A07, TR-A08) separately. The recent onset group performed with 81.2% (SD = 39.2) accuracy in the ObjectWhich condition, with 68.7% (SD = 46.5) in ObjectWho condition, with 62.5% (SD = 48.6) in the SubjectWhich condition, and with 68.7% (SD = 46.5) in the SubjectWho condition. The recent onset group performed significantly better in the Object-Which condition than in the SubjectWhich condition (β = −1.162, SE = 3.713, z = −3.128, p = .009). No other difference returned significant (all ps > .12). The Turkish IWA with more than six months post onset performed better in the ObjectWhich condition (69.0%, SD = 46.3) than in the SubjectWhich (42.2%, SD = 49.5; β = −1.203, SE = 0.238, z = −5.046, p < .001) and SubjectWho (40.4%, SD = 49.2; β = −1.283, SE = 0.239, z = −5.360, p < .001) conditions. Furthermore, these IWA were more accurate responding to the ObjectWho (57.1%, SD = 49.6) than to the SubjectWhich (β = −0.649, SE = 0.229, z = −2.825, p = .024) and SubjectWho conditions (β = −0.729, SE = 0.230, z = −3.160, p = .008). The Turkish IWA with earlier onset did not differ in their responses to the Object-Which and ObjectWho conditions (β = −0.554, SE = 0.237, z = −3.339, p = .008), or in their responses to the SubjectWho and SubjectWhich conditions (β = −0.079, SE = 0.023, z = −0.346, p = .985), however.

The German IWA, on the other hand, performed significantly better in the SubjectWhich condition than in the ObjectWhich (β = 0.943, SE = 0.247, z = 3.814, p < .001), SubjectWho (β = −0.875, SE = 0.246, z = −3.550, p = .002), and ObjectWho conditions (β = 0.913, SE = 0.246, z = 3.697, p = .001). No other comparison proved statistically significant (all ps > .991).

Further post hoc tests confirmed that the Turkish IWA outperformed the German IWA in the ObjectWhich (β = 1.654, SE = 0.482, z = 3.43, p < .001) and the ObjectWho conditions (β = 1.008, SE = 0.380, z = 2.65, p = .008). The German IWA, however, outperformed the Turkish ones in the SubjectWhich condition (β = −0.504, SE = 0.213, z = −2.36, p = .017). The Turkish and German groups did not differ in their responses to the SubjectWho conditions (β = 0.467, SE = 0.604, z = 0.77, p = .433).

### 3.2. Predictors of comprehension failure

The IWA in both language groups showed large individual differences (see Table 4). In order for us to determine which factors contributed to this variation in our participants’ performance, we implemented a RF on each of Turkish and German groups.

Figure 2 demonstrates the variable importance ranking, and Figure 3 shows the best conditional inference tree for our Turkish IWA’s response accuracy.

Our conditional variable importance calculations indicate that the case (nominative vs. accusative) of both wh-elements and non-wh NPs, followed by the gender of the NP associated with the theme role, were among the most informative predictors. Other potential predictors close to the minimum relevance point (shown by dashed red line in Figure 2) contributed little or nothing towards explaining our IWA’s difficulties in comprehending wh-questions. The conditional inference tree for the Turkish data in Figure 3 indicates that the Turkish IWA’s performance in comprehending wh-questions can be broadly predicted by the presence of an accusative case morpheme on the wh-elements (p < .001). According to this classification, the IWA comprehended sentences with accusative marked wh-elements better than sentences containing nominative (i.e., zero-) marked wh-elements.

Figure 4 illustrates the variable importance ranking, and Figure 5 exhibits the best conditional inference tree, for our German IWA’s response accuracy.

For our German IWA, the conditional inference tree in Figure 4 indicates that case ambiguity is a highly informative variable in determining how likely an IWA would fail comprehending our stimulus sentences. The case of both wh-elements and non-wh NPs, as well as the presence of argument-order changing wh-movement proved to be informative factors as well. The conditional inference tree in Figure 5 suggests that the German IWA’s responses were more likely to be accurate when a non-wh NP carried accusative case (i.e., for subject questions) than when it carried nominative case (p = .003; i.e., for object questions). Within the responses to items containing accusative NPs, our inference tree indicates that the German IWA failed to comprehend sentences with ambiguous case cues more often than when case was marked unambiguously (p = .050).
3.3. Summary of results

The findings from our experiments indicate that object which and object who questions are significantly better retained in Turkish IWA than both subject who and subject which questions. The German IWA, by contrast, attained better scores in subject which questions than in all other conditions. Both the Turkish and German IWA’s performance was largely predicted by the presence of accusative case marking. The Turkish IWA tended to perform better for sentences with accusative-marked wh-elements while the German IWA performed better for sentences with accusative-marked non-wh NPs. Importantly, our findings indicate that German IWA perform worse for sentences with ambiguous case marking than for unambiguous case marking.

4. Discussion

Our findings provide insight into the nature of deficits in IWA’s comprehension of wh-questions in both Turkish, a non-wh-movement language, and German, a wh-movement language. We explored the following research questions: (a) whether or not German- and Turkish-speaking IWA differed in their ability to comprehend wh-questions, and (b) which factors best predict their deficits in the comprehension of wh-questions.

With regard to our first research question, the theoretical hypotheses mentioned earlier gave rise to different predictions regarding our IWA’s performance. We followed tradition in dividing these hypotheses into “representational deficit” and “processing deficit” accounts, although to what extent this distinction is in fact warranted is not fully clear. The line of reasoning behind representational deficit account is that IWA’s comprehension difficulties reflect an inability to compute syntactic representations in which fronted constituents are linked to their base positions or “traces” (Grodzinsky, 1995, 2000). As a result, these constituents’ thematic roles cannot be recovered and must be determined in some other way, either by guessing or by applying a default “agent-first” strategy. The inaccessibility of movement traces should affect sentences that involve overt, argument-order-changing movement (such as German object questions) more strongly than sentences in which the canonical argument order is preserved (as
was the case in our Turkish materials, and for German subject questions). From this point of view, Turkish IWA’s comprehension of *wh*-questions might be expected to be largely spared. This expectation was not confirmed by our findings, however. The Turkish IWA clearly had more difficulty comprehending subject *which* and subject *who* questions than both object *which* and object *who* questions, even though neither of these require any overt movement. Assuming that the TDH is relevant only for IWA-speaking *wh*-movement languages, our German data should, however, be more informative in this regard. Following the standard assumption that in German, the derivation of both subject and object questions involves *wh*-raising out of VP, IWA’s comprehension of both question types should in principle be affected by an inability to link *wh*-expressions to their traces. Applying a default “agent-first” strategy should then lead to an incorrect labelling of clause-initial *wh*-expressions in German object questions, causing IWA to have difficulty comprehending object questions. Our German-speaking IWA did indeed show an asymmetrical comprehension deficit in that they performed worse for both object *which* and object *who* questions, both of which involve *wh*-fronting of the object, in comparison to subject *which* questions. However, a general movement deficit as proposed, for example, by Drai and Grodzinsky (2006) cannot by itself account for our finding that the German IWA performed better for subject *which* questions than for both subject and object *who* questions. The German IWA’s poor performance for subject *who* questions, in which there is no argument-order-changing *wh*-movement, is not predicted by the TDH and thus casts doubt on the hypothesis that IWA may be better able to compensate for the loss of subject traces than for the loss of object traces in their sentence representations.

In short, we believe that taken together, our data do not provide much support for a general movement deficit. That said, the presence of non-string-vacuous movement nevertheless seems to be an informative factor in understanding aphasic comprehension difficulties. The outcomes from our random forest models clearly identify *wh*-movement as a relevant factor for the German IWA.

Other representational deficit accounts focus on the notion of intervention. Here the crucial factor for degraded comprehension of object questions is the fact that other sentence material intervenes between the fronted NP in object questions and its trace (e.g., Friedmann et al., 2017; Garraffa & Grillo, 2008; Sheppard et al., 2015). That is, when a *wh*-element is fronted, the presence of intervening words is assumed to render the interpretation of such questions impossible in aphasia. Intervention accounts are similar to the TDH insofar as they also attribute IWA’s poor comprehension of sentences with fronted objects to a deficit in computing movement dependencies. Friedmann et al. (2017) and Friedmann and Shapiro (2003) further proposed that the presence of accusative case marking does not help IWA to recover the moved constituents’ thematic roles. Note that this was what Burchert, De Bleser, and Sonntag (2001) showed for their German IWA’s comprehension of non-canonical sentences, and what Hagiwara and Caplan (1990) showed for Japanese IWA. There are two implications of our data with regard to the presence of intervening sentence material and overt case
Figure 4. Variable importance of potential predictors for German IWA’s (IWA = individuals with non-fluent aphasia) response accuracy for the comprehension of wh-questions. The x-axis indicates variable importance values (higher = more informative). The dashed line indicates minimum relevance. Predictors on the left side of the line contribute nothing towards our understanding of German IWA’s comprehension patterns. NP = noun phrase. [To view this figure in colour, please see the online version of this Journal.]

Figure 5. The best fitting conditional binary inference tree for the German IWA’s (IWA = individuals with non-fluent aphasia) comprehension of wh-questions (total number of classified observations = 576). Nodes 1 and 3 indicate significant branching points (significance shown in p-value). Proportions of participants’ accurate responses to the picture-pointing task (1 = accurate response) are indicated on the right of each box-plot. The numbers in parentheses above each plot show the number of classified observations (out of 576). Case non-wh NP = case of non-wh noun phrase (nominative vs. accusative); case ambiguity = whether or not the sentence stimulus contains an ambiguous case cue.
marking not being helpful during IWA’s comprehension. Regarding the first point, recall that in German object questions the subject NP intervenes between the fronted object and its trace(s), whilst there is no intervening NP in German subject questions. Our German IWA’s comprehension difficulty of both the object which and who questions is clearly consistent with the intervention hypothesis. However, the German IWA’s poor performance for subject who relative to subject which questions cannot be accounted for by the intervention account. Regarding the second point—that is, the claim that the availability of accusative case marking in non-canonical sentences to signal thematic roles is unhelpful for IWA—our data seem to support the findings from previous studies (Burchert et al., 2001; Friedmann et al., 2017; Friedmann & Shapiro, 2003; Hagiwara & Caplan, 1990). The only conditions where word-order-changing wh-movement applied were German object which and who questions in our experiments. The German IWA found both kinds of object question hard to comprehend. Thus, accusative marking on fronted wh-expressions did not seem to have helped our German IWA. Having said this, recall that German case marking is ambiguous for feminine NPs and that case ambiguity was identified as a relevant factor in our RF analysis. We return to this issue below.

The second type of account that we considered here were processing accounts, according to which IWA do not necessarily have impaired syntactic representations but may show delayed sentence processing. Of particular relevance to the current study is the discourse-linking hypothesis as this makes specific predictions about IWA’s interpretation of wh-questions (e.g., Avrutin, 2000, 2006; Bos et al., 2014; Hickok & Avrutin, 1996; Nerantzini et al., 2014; Salis & Edwards, 2008). The starting assumption here is that which X phrases refer to a particular entity in the preceding discourse and thus are “discourse-linked” whereas wh-pronouns are not. Processing discourse-linked entities is assumed to incur greater processing cost than processing non-discourse-linked elements due to the need to access and integrate discourse-level information in the former case. Therefore, according to the discourse-linking hypothesis, which questions are expected to be more challenging for IWA than who questions, irrespective of movement. From this perspective, we may have expected both our Turkish and German IWA to be equally impaired in comprehending which questions in comparison to who questions. This is, however, not what we saw in our data. We found who questions to be considerably affected in our German IWA, which is consistent with previous studies that showed who questions to be impaired in aphasia (e.g., Cho-Reyes & Thompson, 2012; Neuhaus & Penke, 2008; Salis & Edwards, 2008). Moreover, the German IWA found subject which questions easier to understand than object which questions, whereas the Turkish IWA found object which questions easier than subject which ones. The different directions of impairment for Turkish and German which questions are difficult to account for from the point of view of the discourse-linking hypothesis.

Based on the current data we are unable to assess whether or not our IWA encountered difficulties comprehending wh-questions due to slowed or weakened processing as we only measured their offline responses. Many studies have shown IWA’s online sentence processing patterns to be similar to those of non-impaired controls except that processing may be somewhat delayed, whilst the IWA’s offline responses indicate comprehension failure (Dickey & Thompson, 2009; Hanne et al., 2016). Future research examining moment-by-moment processing in a non-wh-movement language would be informative.

The German IWA’s relatively spared ability to process subject which questions compared to object ones is broadly consistent with the results from earlier studies on German IWA (Hanne et al., 2016; Neuhaus & Penke, 2008). Interestingly, however, our German- and Turkish-speaking groups of IWA patterned dissimilarly in their comprehension of subject which and object which questions. But what, then, makes subject which questions hard for Turkish IWA but relatively easy for German ones, and what makes object which questions difficult for German IWA but easier for Turkish ones? We believe this has to do with the transparency and ambiguity of case marking in the two languages. The random forest machine learning algorithm we ran on the Turkish data broadly predicted that Turkish IWA’s response accuracy was determined by the presence of accusative case marking on the wh-expressions (see Figure 3). That is, Turkish IWA are likely to perform more accurately when there is an accusative case marker on the wh-expression than when the wh-expression carries nominative case. For the German IWA the algorithm showed that their responses are more likely to be accurate when a non-wh NP carried
accusative than when these carried nominative—that is, for canonical subject questions. However, the conditional inference tree (Figure 5) also shows that the presence of ambiguous case marking (that is, of the feminine article die or the wh-determiner welche, which are the same in nominative and accusative case) significantly affected sentence interpretation ability in German aphasia. German IWA’s ability to interpret unambiguous case marking is in line with the findings reported by Hanne et al. (2015).

In contrast to case, the gender features of the NPs involved did not influence our IWA’s comprehension in any measurable way. According to our variable importance calculations, gender was ranked as the least important factor (Figure 4). In other words, although case ambiguity is specific to feminine NPs in German, what influenced IWA’s comprehension ability was the presence of ambiguous case forms, not whether the NP carrying the theme role referred to a feminine entity or not.

The picture seems to be less complicated in Turkish than in German. Turkish nominatives are zero-marked (i.e., there is no overt morpheme), and thus it seems that the Turkish IWA relied on the presence of transparent accusative marking on wh-elements to interpret those questions. This conclusion is largely compatible with the findings reported by MacWhinney et al. (1991) and Yarbay-Duman et al. (2011) for Turkish IWA, and with those of Kljajevic and Murasugi (2010) for Croatian, another relatively free word-order language. For instance, Yarbay-Duman and colleagues found that in Turkish passive constructions, where both the agent and the theme arguments receive nominative case (i.e., zero-marking), Turkish IWA encountered greater difficulties than when the case assignment followed the canonical order of agent (nominative) and theme (accusative). Our study thus provides converging evidence that both Turkish and German IWA’s interpretation of questions is guided by the presence of unambiguous and transparent case marking (in contrast to Burchert et al., 2003; Friedmann et al., 2017). Higher cue reliability in Turkish—that is, the lack of ambiguous accusative forms—seems to have worked in favour of our Turkish IWA, who performed better for wh-questions containing transparent accusative-marked wh-elements than for zero-marked nominatives.

Finally, we would like to draw attention to the importance of machine learning algorithms in exploring and analysing data for subtle differences. Aphasia research has constantly had to deal with large individual variability. Most studies on wh-questions that we reviewed above contained groups of IWA in whom severity of cognitive deficits, response accuracy patterns, and even the aphasia types were not similar across individuals. Measuring group means in such groups seems far from helpful and may even be misleading. Using logistic mixed-effect regression models we are now able to incorporate particular individuals and items as random factors. However, a problem with the use of logistic regression models for analysing aphasia data is that they may not be able to accommodate interactions between too many factors, especially when these factors correlate with each other. Thus a good option here seems using mixed-effect regression models with random forest machine learning algorithms, as suggested by Tagliamonte and Baayen (2012). Another advantage of using random forest algorithms is that they allow us to determine whether, and to what extent, several linguistic features are informative.

In conclusion, the results from the current study show that IWA speaking Turkish, a non-wh-movement language, and German, a wh-movement language, differed in their comprehension ability of different types of wh-question. We argued that IWA’s sentence comprehension difficulties cannot be accounted for by attributing these difficulties simply to the presence of argument-order-changing wh-movement or to difficulties integrating discourse level information. Instead, our findings suggest that sentence comprehension processes in Turkish and German non-fluent aphasia rely to a large extent on unambiguous and transparent case marking that indicate thematic role assignments.

Notes

1. Objects may, however, be scrambled or topicalized, yielding alternative ordering patterns. We do not discuss these phenomena any further here as they were not investigated in the current study.
2. For ease of exposition we have omitted indicating possible additional movements through an extended VP or split-CP system.
3. See Özsoy (2009) for an overview of accounts on scrambling conditions where wh-movement can be licensed to fulfil a number of pragmatic functions. These particular conditions are out of the scope of the current investigation.
4. Analyses in Kornfilt (2003) indicate that the unmarked (nominative) objects in Turkish necessarily appear in the immediate pre-verbal position while accusative...
marked ones are free to scramble to any other location within the sentence. In this study, we explore the SOV alignment where scrambling does not apply.

5. The verbs chosen for the experiments are controlled for their lemma frequencies (per million tokens), which were retrieved from the DLexDB database for German (Heister et al., 2011) and the TS Corpus for Turkish (Sezer & Sezer, 2013).

6. Severity was calculated for each IWA by computing the sum of their scores in the aphasia assessment subsections (including sentence comprehension, word comprehension, word repetition, token test, and naming). The raw scores in each subsection were first converted to percentages and then summed prior to entering them into the analyses.

7. Since the Turkish IWA tended to have a more recent onset of aphasia than the German IWA, we included post-onset time as a factor in the analyses, based on the assumption that a recent onset might lead to more severe aphasic syndromes. Post-onset time yielded no significant effects in the main mixed-effects regression analysis, however. Additionally, correlations between post-onset and aphasia severity turned out to be weak (Spearman $r_s = -0.28, n = 1,440$). Therefore, any condition differences in our IWA group seem relatively unlikely to be caused by a more recent onset of aphasia.

Acknowledgements

We are grateful to the Cognitive Neuropsychology editors and reviewers for their positive and constructive feedback on our earlier manuscript. We thank İknur Maviş, Figen Schultz-Ünsal, and our research assistant Kübra Güleç for their help during participant recruitment. We also thank our participants for taking part in our experiments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The study reported here has been funded by an Alexander von Humboldt-Stiftung Professorship awarded to Harald Clahsen, Potsdam Research Institute for Multilingualism, University of Potsdam.

References


In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), Language and the brain: Representation and processing (pp. 295–313). New York: Academic Press.


Appendices

Appendix A. Lemma frequencies of verbs (per million tokens) used in constructing the sentence stimuli of our Turkish and German experiments.

<table>
<thead>
<tr>
<th>Verb used</th>
<th>Lemma frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>to smack</td>
<td>1.392</td>
</tr>
<tr>
<td>to kick</td>
<td>1.052</td>
</tr>
<tr>
<td>to kiss</td>
<td>20.000</td>
</tr>
<tr>
<td>to pull</td>
<td>513.211</td>
</tr>
<tr>
<td>to push</td>
<td>29.314</td>
</tr>
<tr>
<td>to chase</td>
<td>6.366</td>
</tr>
<tr>
<td>to bite</td>
<td>4.450</td>
</tr>
<tr>
<td>to tickle</td>
<td>626</td>
</tr>
<tr>
<td>to poke</td>
<td>894</td>
</tr>
<tr>
<td>to stroke</td>
<td>417.5</td>
</tr>
<tr>
<td>to choke</td>
<td>15.982</td>
</tr>
<tr>
<td>to pinch</td>
<td>207</td>
</tr>
</tbody>
</table>

Appendix B. List of sentences used in the experiments

Object which questions
1. **Turkish test**: Adam/kadın adamı/adamını tokatıyor?
   **German test**: Welchen Mann/Welche Frau schlägt der Mann/die Frau? **Eng. Trans.**: Which man/woman is the man/woman smacking?
2. Adam/kadın adamı/adamını tekemeliyor?
Welchen Mann/Welche Frau tritt der Mann/die Frau? Which man/woman is the man/woman kicking?
3. Adam/kadın adamı/adamını öpüyor?
Welchen Mann/Welche Frau küsst der Mann/die Frau? Which man/woman is the man/woman kissing?
4. Adam/kadın adamı/adamını çekiyor?
Welchen Mann/Welche Frau zieht der Mann/die Frau? Which man/woman is the man/woman pulling?
5. Adam/kadın adamı/adamını itiyor?
Welchen Mann/Welche Frau stößt der Mann/die Frau? Which man/woman is the man/woman pushing?
6. Adam/kadın adamı/adamını kovalıyor?
Welchen Mann/Welche Frau kitzelt der Mann/die Frau? Which man/woman is the man/woman tickling?
7. Adam/kadın adamı/adamını sıçıyor?
Welchen Mann/Welche Frau fegt der Mann/die Frau? Which man/woman is the man/woman choking?
8. Adam/kadın adamı/adamını gidyokiyor?
Welchen Mann/Welche Frau küsztelt der Mann/die Frau? Which man/woman is the man/woman biting?
9. Adam/kadın adamı/adamını diüttür?
Welchen Mann/Welche Frau kiekt der Mann/die Frau? Which man/woman is the man/woman poking?
10. Adam/kadın adamı/adamını okşuyor?
Welchen Mann/Welche Frau streicht der Mann/die Frau? Which man/woman is the man/woman stroking?
11. Adam/kadın adamı/adamını bojuyor?
Welchen Mann/Welche Frau würtzt der Mann/die Frau? Which man/woman is the man/woman choking?
12. Adam/kadın adamı/adamını çımdikliyor?
Welchen Mann/Welche Frau kneift der Mann/die Frau? Which man/woman is the man/woman pinching?

Object who questions
13. **Turkish test**: Adam/kadın kimi tokatıyor?
    **German test**: Wen schlägt der Mann/die Frau? **Eng. Trans.**: Who is the man/woman smacking?
14. Adam/kadın kimi tekemeliyor?
    Wen tritt der Mann/die Frau? Who is the man/woman kicking?
15. Adam/kadın kimi öpüyor?
    Wen küsst der Mann/die Frau? Who is the man/woman kissing?
16. Adam/kadın kimi çekiyor?
    Wen zieht der Mann/die Frau? Who is the man/woman pulling?
17. Adam/kadın kimi itiyor?
    Wen stößt der Mann/die Frau? Who is the man/woman pushing?
Subject who questions

25. **Turkish test:** Hangi adam/kadın adamı/kadını tokatlıyor?
    German test: Welcher Mann/Welche Frau schlägt den Mann/die Frau?
    Eng. Trans.: Who is smacking the man/woman?

26. Hangi adam/kadın adamı/kadını tekmeleyiyor?
    Welcher Mann/Welche Frau tritt den Mann/die Frau?
    Which man/woman is kicking the man/woman?

27. Hangi adam/kadın adamı/kadını öpüyor?
    Welcher Mann/Welche Frau küsst den Mann/die Frau?
    Which man/woman is kissing the man/woman?

28. Hangi adam/kadın adamı/kadını çekiyor?
    Welcher Mann/Welche Frau zieht den Mann/die Frau?
    Which man/woman is pulling the man/woman?

29. Hangi adam/kadın adamı/kadını itiyor?
    Welcher Mann/Welche Frau stößt den Mann/die Frau?
    Which man/woman is pushing the man/woman?

30. Hangi adam/kadın adamı/kadını kovalıyor?
    Welcher Mann/Welche Frau jagt den Mann/die Frau?
    Which man/woman is chasing the man/woman?

31. Hangi adam/kadın adamı/kadını ısıyor?
    Welcher Mann/Welche Frau beißt den Mann/die Frau?
    Which man/woman is biting the man/woman?

32. Hangi adam/kadın adamı/kadını gıdikleyor?
    Welcher Mann/Welche Frau kitzelt den Mann/die Frau?
    Which man/woman is tickling the man/woman?

33. Hangi adam/kadın adamı/kadını dördürüyor?
    Welcher Mann/Welche Frau streichelt den Mann/die Frau?
    Which man/woman is stroking the man/woman?

34. Hangi adam/kadın adamı/kadını okuyor?
    Welcher Mann/Welche Frau pieskt den Mann/die Frau?
    Which man/woman is poking the man/woman?

35. Hangi adam/kadın adamı/kadını boğuyor?
    Welcher Mann/Welche Frau kitzelt den Mann/die Frau?
    Which man/woman is tickling the man/woman?

Appendix C. Example codes for running the random forest algorithm in R

- A random forest with unbiased decision trees is computed with:
  ```r
  # fit <- cforest(as.factor(accuracy) ~ ., data = accuracy, control = cforest_unbiased (mtry = 3 ntree = 2000))
  ```
- Assessment of conditional variable importance is computed with:
  ```r
  # varimp(fit, conditional = TRUE)
  ```
- A conditional inference tree is plotted with:
  ```r
  # Tree <- ctree(as.factor(accuracy) ~ ., data = accuracy, control = cforest_unbiased); plot(Tree)
  ```