Humanwissenschaftliche Fakultät

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

COGNITIVE NEUROPSYCHOLOGY, 2017
VOL. 34, NO. 5, 312–331
https://doi.org/10.1080/02643294.2017.1394284

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Predicting the sources of impaired *wh*-question comprehension in non-fluent aphasia: A cross-linguistic machine learning study on Turkish and German

Seçkin Arslan[a], Eren Gür[b] and Claudia Felser[a]

[a]Potsdam Research Institute for Multilingualism, University of Potsdam, Potsdam, Germany; [b]Department 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.[1] 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$_{subj}$ $[_{C'}$ $[_{C}$ küsst$_{i}$] $[_{IP}$ t$_{subj'}$ . . . $[_{VP}$ t$_{subj}$ $[_{V'}$ den Mann $[_{V}$ t$_{i}$]]]]]] ?

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′).

(2) Wen küsst der Mann ?
who.$_{ACC}$ kisses the.$_{NOM}$ man
Who is the man kissing?

(2′) [$_{CP}$ Wen$_{obj}$ [$_{C'}$ [$_{C}$ küsst$_{i}$] [$_{IP}$ der Mann$_{subj}$ . . . [$_{VP}$ t$_{subj}$ [$_{V'}$ t$_{obj}$ [$_{V}$ t$_{k}$]]]]]] ?

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; Özsoy, 2009).[3] See Examples 3 and 4 for illustration (PRES.PROG. = present progressive).

(3) Kim adamı öpüyor?
who.$_{NOM}$ man.$_{ACC}$ kiss.$_{PRES.PROG.}$
Who is kissing the man?

(4) Adam kimi öpüyor?
man.$_{NOM}$ who.$_{ACC}$ kiss.$_{PRES.PROG.}$
Who is the man kissing?

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.[4] 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, Walenski, 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; Nerantzini, 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, Nerantzini 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.

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. 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, Bamyacı, & Bastiaanse, 2016; Aydin, 2007; MacWhinney, Osmán-Sági, & Slobin, 1991; Yarbay-Duman, Altınok, Özgirgin, & 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 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 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.

Table 1. Demographic and medical details of the participants with aphasia.

Participant	Gender	Education (years)	Occupation	Age	Hand.	MPO	Aetiology	Lesion location
DE-A01	M	12	Worker	54	R	84	Cardio-embolic stroke	Multiple infarcts over left middle watershed area
DE-A02	F	12	Retired	72	R	60	Cardio-embolic stroke	Left fronto-temporal
DE-A03	M	10	Worker	71	R	48	Cardio-embolic stroke	Left temporal-parietal
DE-A04	F	10	Nurse	51	R	180	Atherosclerotic stroke	Left fronto-temporal
DE-A05	M	14	Worker	44	R	132	Atherosclerotic stroke	Left medial and posterior regions
DE-A06	M	13	Information officer	64	L	264	Thalamic putaminal lober haematoma	Right Fronto-parietal regions
TR-A01	M	12	Worker	51	R	48	Cardio-embolic stroke	Left fronto-temporal
TR-A02	M	10	Retired	66	R	180	Atherosclerotic stroke	Left cerebellar, fronto-temporal, parietal
TR-A03	M	10	Worker	65	R	14	Hypertensive subdural & epidural hematoma	Left temporal-parietal
TR-A04	M	12	Retired	69	L	13	Cardio-embolic stroke	Right fronto-temporal
TR-A05	F	8	Unemployed	61	R	11	Atherosclerotic stroke	Left temporal-parietal
TR-A06	M	5	Retired	56	R	3	Atherosclerotic stroke	Left fronto-temporal, parietal
TR-A07	M	12	Accountant	57	R	22	Atherosclerotic stroke	Left fronto-temporal
TR-A08	M	9	Driver	56	R	9	Cardio-embolic stroke	Left fronto-temporal
TR-A09	M	8	Retired	73	R	5	Cardio-embolic stroke + haematoma	Left frontal infarct, parietal-occipital hypodense
TR-A10	M	8	Retired	63	R	3	Atherosclerotic stroke	Left fronto-temporal
TR-A11	F	6	Unemployed	74	R	3	Atherosclerotic stroke	Left fronto-temporal, parietal

Note: DE = German agrammatic speaker; TR = Turkish agrammatic speaker; M = male; F = female; Hand. = handedness; MPO = months post onset. Aetiology and lesion location information, based on computed tomography (CT) scans, are taken from the participants' medical reports. Note that DE-A06 and TR-A04 are left-handers with right-brain lesions.

Table 2. The participants' raw scores from aphasia language assessments and maximum possible score in the relevant section.

Participant	Token test	Comprehension AAT/ADD		Repetition AAT/ADD		Naming AAT/ADD
		Word	Sentence	Word	Sentence	
DE-A01	20/50 (40)	45/60 (75)	50/60 (83.3)	116/120 (96.6)	20/30 (66.6)	90/120 (75)
DE-A02	12/50 (24)	24/60 (60)	15/60 (25)	44/120 (36.6)	6/30 (20)	24/120 (20)
DE-A03	22/50 (44)	49/60 (81.5)	37/60 (61.6)	78/120 (65)	4/30 (13.3)	59/120 (49.1)
DE-A04	35/50 (70)	42/60 (70)	43/60 (71.6)	101/120 (84.1)	3/30 (10)	68/120 (56.6)
DE-A05	3/50 (6)	36/60 (60)	23/60 (38.3)	45/120 (37.5)	1/30 (3.3)	105/120 (87.5)
DE-A06	35/50 (70)	59/60 (83.3)	40/60 (66.66)	99/120 (82.5)	12/30 (40)	120/85 (70.8)
TR-A01	6/50 (12)	3/32 (9.3)	4/34 (11.7)	7/20 (35)	NA	2/20 (8.3)
TR-A02	10/50 (20)	12/32 (37.5)	24/34 (41.1)	9/20 (45)	NA	15/24 (62.5)
TR-A03	42/50 (84)	24/32 (75)	22/34 (64.7)	19/20 (95)	NA	6/24 (25)
TR-A04	41/50 (82)	20/32 (60.5)	21/34 (61.7)	19/20 (95)	NA	22/24 (91.6)
TR-A05	38/50 (76)	10/32 (31.2)	10/34 (29.4)	15/20 (75)	NA	22/24 (91.6)
TR-A06	41/50 (82)	22/32 (68.7)	24/34 (70.5)	19/20 (95)	NA	21/24 (87.5)
TR-A07	13/50 (26)	15/32 (46.8)	14/34 (35.2)	12/20 (60)	NA	16/24 (80)
TR-A08	18/50 (36)	3/32 (9.3)	4/34 (11.7)	10/20 (50)	NA	2/24 (8.3)
TR-A09	8/50 (16)	11/32 (34.3)	12/34 (35.2)	8/20 (40)	NA	13/24 (54.1)
TR-A10	35/50 (70)	14/32 (43.7)	13/34 (38.2)	15/20 (75)	NA	16/24 (66.6)
TR-A11	16/50 (32)	24/32 (75)	22/34 (64.7)	19/20 (95)	NA	15/24 (62.5)

Note: Percentages are given in parentheses. DE = German agrammatic speaker; TR = Turkish agrammatic speaker. The diagnostic assessment scores were obtained by using the Aachen Aphasia Test (AAT; Huber et al., 1983) for German, and the Aphasia Language Assessment (ADD; Maviş & Toğram, 2009) for Turkish; the absolute scores were converted to percentages for ease of comparability. The Token Test (De Renzi & Faglioni, 1978) was administered in full to each participant. Note that the ADD for Turkish contains one repetition section where the majority of the items are single words (one or two sentences). Therefore, repetition scores for word level are reported only. Also note that word level comprehension scores from the ADD show individuals' performance on a "single word" comprehension task (i.e., requiring the individuals to identify different semantic categories). However, the sentence level comprehension part contains sections with short commands, yes/no questions, and simple to complex affirmative sentences in a picture-matching design.

In addition, two groups of German ($n = 17$, $M_{age} = 61.82$ years, range = 52–68 years) and Turkish ($n = 13$, $M_{age} = 62.38$ 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*,

Table 3. Example sentence stimuli used in the experiments.

Condition	Sentence stimuli				
1. ObjectWhich	TR: Kadın	hangi	adamı	çekiyor?	
	woman	which	man.ACC	pull.PRES.PROG.	
	DE: Welchen	Mann	zieht	die Frau?	
	which.ACC	man	pulls	the woman	
	ENG: Which man is the woman pulling?				
2. ObjectWho	TR: Kadın	kimi	çekiyor?		
	woman	who.ACC	pull.PRES.PROG.		
	DE: Wen	zieht	die Frau?		
	who.ACC	pulls	the woman		
	ENG: Who is woman pulling?				
3. SubjectWhich	TR: Hangi	kadın	adamı	çekiyor?	
	which	woman.NOM	man.ACC	pull.PRES.PROG.	
	DE: Welche	Frau	zieht	den	Mann?
	which	woman	pulls	the.ACC	man
	ENG: Which woman is pulling the man?				
4. SubjectWho	TR: Kim	adamı	çekiyor?		
	who	man.ACC	pull		
	DE: Wer	zieht	den	Mann?	
	Who.NOM	pulls	the.ACC	man?	
	ENG: Who is pulling the man?				

Note: TR = Turkish sentence stimuli; DE = German sentence stimuli; ENG = English translation; ACC = accusative; NOM = nominative.

SubjectWhich, and *SubjectWho* questions. Table 3 presents example sentence stimulus sets.

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*?).[5] 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 (ObjectWhich 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. In the other two versions, a female actor appeared in the middle and two male actors on her sides, again 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 ObjectWho 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

Figure 1. An example visual stimulus used in the experiments (the action "to push" is depicted). [To view this figure in colour, please see the online version of this Journal.]

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: $ß = 4.3689$, $SE = 0.361$, $z = 12.103$, $p < .001$). 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 4. Accuracy proportions of the Turkish and German IWA's comprehension of *wh*-questions.

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

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

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[6] 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.[7] 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 ObjectWhich 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 SubjectWho 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 5. Outputs from the main generalized mixed-effects regression model for the aphasia data.

	Estimate (ß)	*SE*	*z*	*p*
(Intercept)	−2.250	3.839	−5.862	<.001
Post onset	2.954e	1.430e	0.021	.983
Severity	2.961	4.084e	7.251	<.001
Condition (ObjectWho)	6.993	3.909	1.789	.073
Condition (SubjectWhich)	1.410	3.238	4.356	<.001
Condition (SubjectWho)	6.682	3.013	2.218	.026
Language (Turkish)	1.792	2.999	5.976	<.001
Condition (ObjectWho) × Language	−1.333	4.395	−3.033	.002
Condition (SubjectWhich) × Language	−2.598	3.810	−6.819	<.001
Condition (SubjectWho) × Language	−1.803	3.618	−4.984	<.001

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 ObjectWhich condition than in the SubjectWhich condition ($\beta = -1.162$, $SE = 3.713$, $z = -3.128$, $p = .009$). 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%, $SD = 49.5$; $\beta = -1.203$, $SE = 0.238$, $z = -5.046$, $p < .001$) and SubjectWho (40.4%, $SD = 49.2$; ß $= -1.283$, $SE = 0.239$, $z = -5.360$, $p < .001$) conditions. Furthermore, these IWA were more accurate responding to the ObjectWho (57.1%, $SD = 49.6$) than to the SubjectWhich (ß $= -0.649$, $SE = 0.229$, $z = -2.825$, $p = .024$) and SubjectWho conditions (ß $= -0.729$, $SE = 0.230$, $z = -3.160$, $p = .008$). The Turkish IWA with earlier onset did not differ in their responses to the ObjectWhich and ObjectWho conditions ($\beta = -0.554$, $SE = 0.237$, $z = -3.339$, $p = .08$), or in their responses to the SubjectWho and SubjectWhich conditions (ß $= -0.079$, $SE = 0.23$, $z = -0.346$, $p = .985$), however.

The German IWA, on the other hand, performed significantly better in the SubjectWhich condition than in the ObjectWhich ($\beta = 0.943$, $SE = 0.247$, $z = 3.814$, $p < .001$), SubjectWho ($\beta = -0.875$, $SE = 0.246$, $z = -3.550$, $p = .002$), and ObjectWho conditions ($\beta = 0.913$, $SE = 0.246$, $z = 3.697$, $p = .001$). 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 ($\beta = 1.654$, $SE = 0.482$, $z = 3.43$, $p < .001$) and the ObjectWho conditions ($\beta = 1.008$, $SE = 0.380$, $z = 2.65$, $p = .008$). The German IWA, however, outperformed the Turkish ones in the SubjectWhich condition ($\beta = -0.504$, $SE = 0.213$, $z = -2.36$, $p = .017$). The Turkish and German groups did not differ in their responses to the SubjectWho conditions ($\beta = 0.467$, $SE = 0.604$, $z = 0.77$, $p = .433$).

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 ($p = .050$).

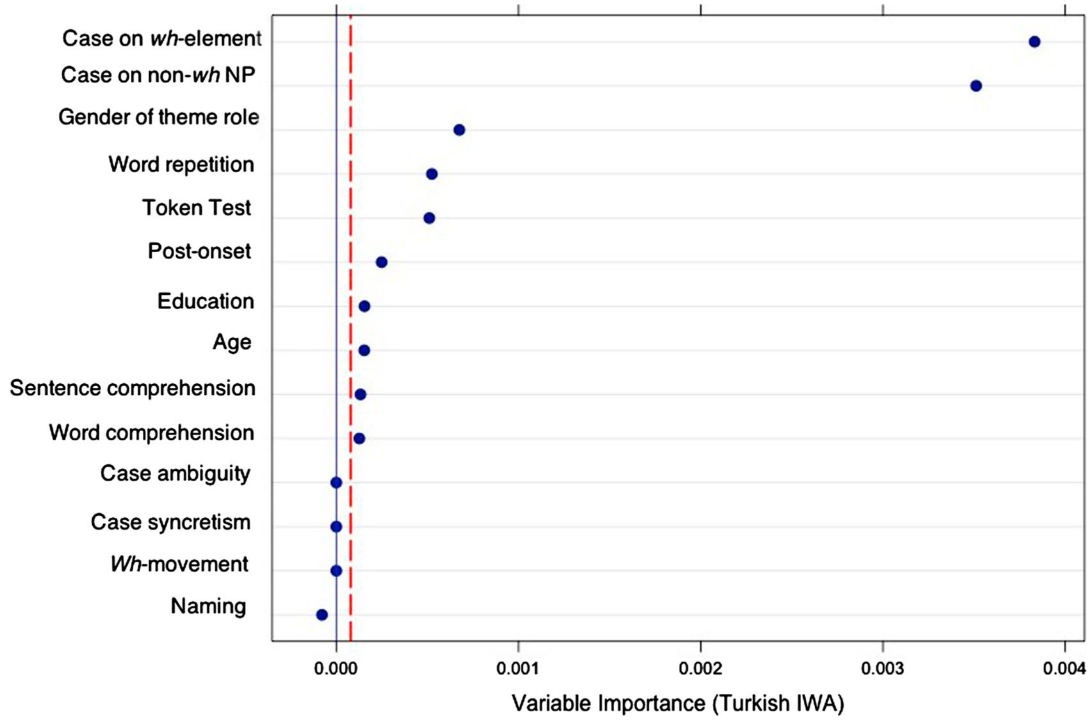

Figure 2. Variable importance of potential predictors for Turkish IWA's (IWA = individuals with non-fluent aphasia) response accuracy for the comprehension of *wh*-questions. The *x*-axis indicates variable importance values (higher = more informative). The dashed line indicates minimum relevance. Predictors on the left side of the line contribute nothing towards our understanding of Turkish IWA's comprehension patterns. NP = noun phrase. [To view this figure in colour, please see the online version of this Journal.]

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

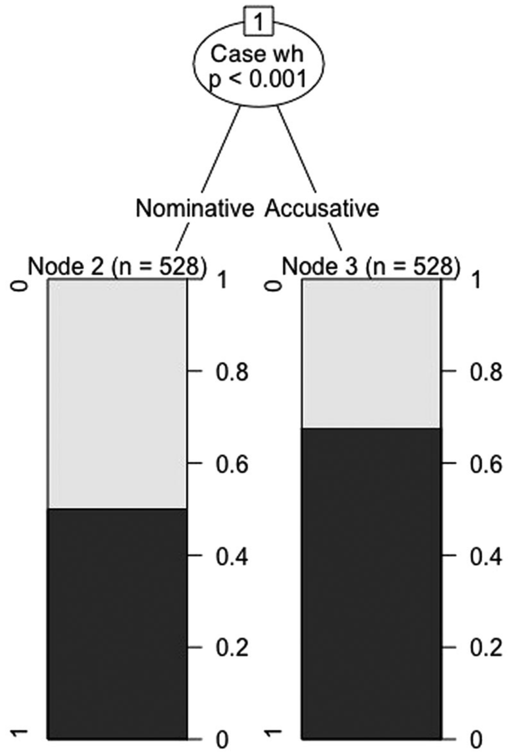

Figure 3. The best conditional binary inference tree for the Turkish IWA's (IWA = individuals with non-fluent aphasia) comprehension of *wh*-questions (total number of classified observations = 1056). Node 1 indicates the significant branching point. Box-plots illustrate Turkish IWA's response accuracy (higher black = more accurate response). The numbers in parentheses above each box-plot indicate the number of observations. Case wh = case of wh-elements (nominative vs. accusative).

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

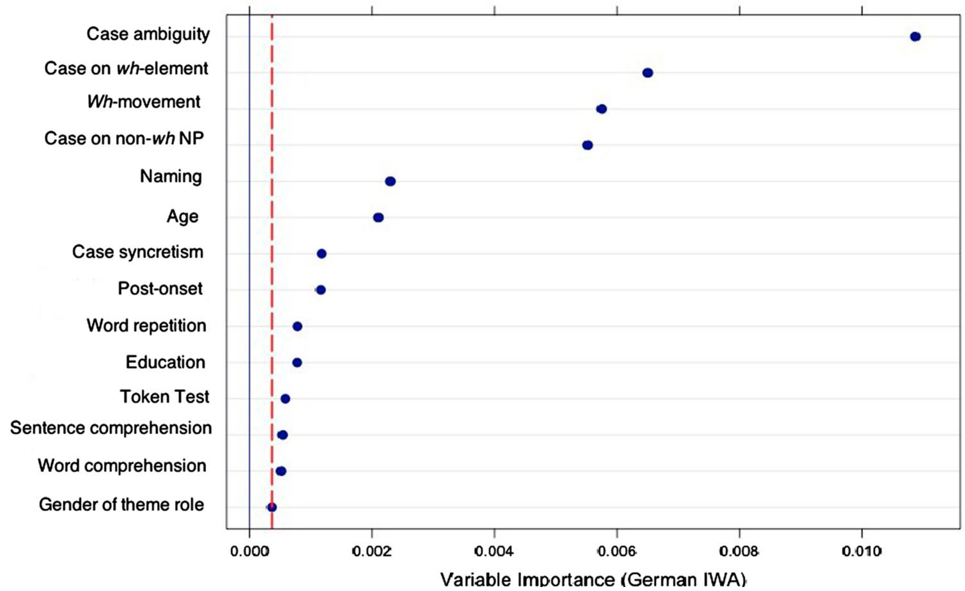

Figure 4. Variable importance of potential predictors for German IWA's (IWA = individuals with non-fluent aphasia) response accuracy for the comprehension of *wh*-questions. The *x*-axis indicates variable importance values (higher = more informative). The dashed line indicates minimum relevance. Predictors 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.]

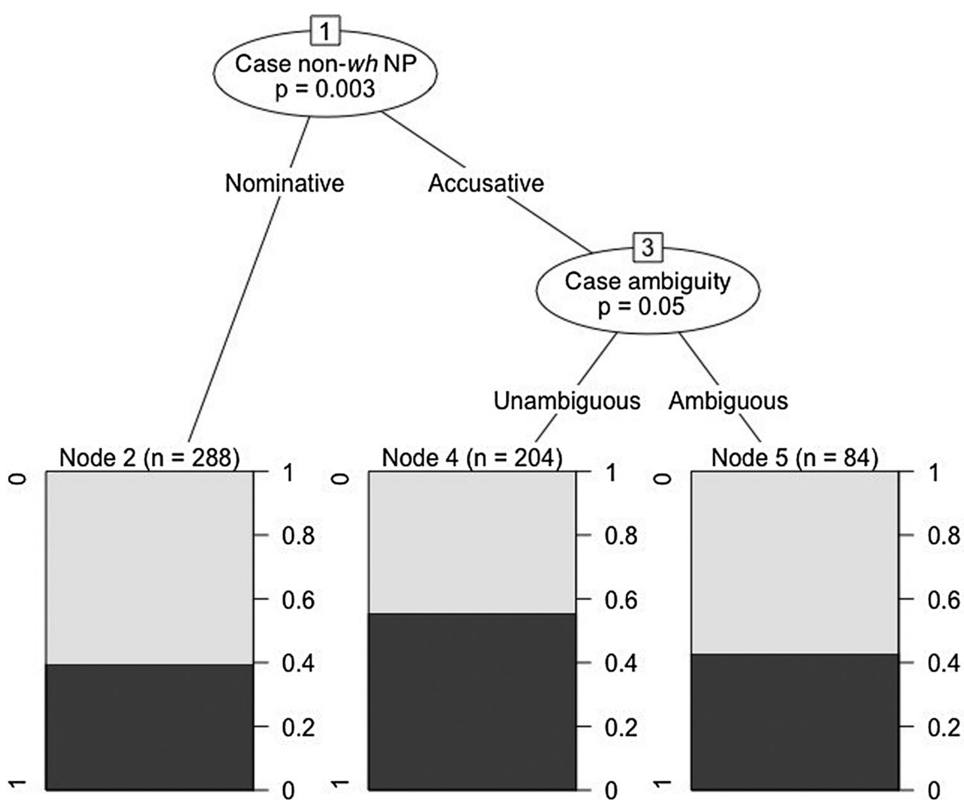

Figure 5. The best fitting conditional binary inference tree for the German IWA's (IWA = individuals with non-fluent aphasia) comprehension of *wh*-questions (total number of classified observations = 576). Nodes 1 and 3 indicate significant branching points (significance shown in *p*-value). Proportions 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 $r_s = -.28$, $n = 1{,}440$). 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 İlknur 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

Akar, D. (1990). *Wh-questions in Turkish* (*M.A. Thesis*). Boğaziçi University.

Arslan, S., Bamyacı, E., & Bastiaanse, R. (2016). A characterization of verb use in Turkish agrammatic narrative speech. *Clinical Linguistics & Phonetics*, *30*(6), 449–469. doi:10.3109/02699206.2016.1144224

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), *Language and the brain: Representation and processing* (pp. 295–313). New York: Academic Press.

Avrutin, S. (2006). Weak syntax. In Y. Grodzinsky, & K. Amunts (Eds.), *Broca's region* (pp. 49–62). Oxford: Oxford University Press.

Aydin, Ö. (2007). The comprehension of turkish relative clauses in second language acquisition and agrammatism. *Applied Psycholinguistics*, *28*(02), 295–315. doi:10.1017/S0142716407070154

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. doi:10.1016/j.jml.2007.12.005

Bach, E. (1962). The order of elements in a transformational grammar of German. *Language*, *38*(3), 263–269. doi:10.2307/410785

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Blumstein, S. E., Byma, G., Kurowski, K., Hourihan, J., Brown, T., & Hutchinson, A. (1998). On-line processing of filler–gap constructions in aphasia. *Brain and Language*, *61*(2), 149–168. doi:10.1006/brln.1997.1839

Bos, L. S., Dragoy, O., Avrutin, S., Iskra, E., & Bastiaanse, R. (2014). Understanding discourse-linked elements in aphasia: A threefold study in Russian. *Neuropsychologia*, *57*, 20–28. doi:10.1016/j.neuropsychologia.2014.02.017

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Burchert, F., De Bleser, R., & Sonntag, K. (2001). Does case make the difference? Agrammatic comprehension of case-inflected sentences in German. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *37*(5), 700–702. doi:10.1016/S0010-9452(08)70618-X

Burchert, F., De Bleser, R., & Sonntag, K. (2003). Does morphology make the difference? Agrammatic sentence comprehension in German. *Brain and Language*, *87*(2), 323–342. doi:10.1016/S0093-934X(03)00132-9

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*(4), 572–582. doi:10.1016/0093-934X(76)90048-1

Chomsky, N. (1981). *Lectures on binding and government*. Dordrecht: Foris.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern assessment of verbs and sentences (NAVS). *Aphasiology*, *26*(10), 1250–1277. doi:10.1080/02687038.2012.693584

de Aguiar, V., Bastiaanse, R., & Miceli, G. (2016). Improving production of treated and untreated verbs in aphasia: A meta-analysis. *Frontiers in Human Neuroscience*, *10*, 89. doi:10.3389/fnhum.2016.00468

den Besten, H. (1983). On the interaction of root transformations and lexical deletive rules. In W. Abraham (Ed.), *On the formal syntax of the westgermania: Papers from the 3rd Groningen grammar talks, Groningen, January 1981* (pp. 47–131). Amsterdam/Philadelphia: John Benjamins Publishing Co.

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the token test. *Cortex*, *14*(1), 41–49. doi:10.1016/S0010-9452(78)80006-9

Dickey, M. W., Choy, J. J., & Thompson, C. K. (2007). Real-time comprehension of wh- movement in aphasia: Evidence from eyetracking while listening. *Brain and Language*, *100*(1), 1–22. doi:10.1016/j.bandl.2006.06.004

Dickey, M. W., & Thompson, C. K. (2009). Automatic processing of wh- and NP-movement in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics*, *22*(6), 563–583. doi:10.1016/j.jneuroling.2009.06.004

Drai, D., & Grodzinsky, Y. (2006). A new empirical angle on the variability debate: Quantitative neurosyntactic analyses of a large data set from Broca's aphasia. *Brain and Language*, *96*(2), 117–128. doi:10.1016/j.bandl.2004.10.016

Friedmann, N., Rizzi, L., & Belletti, A. (2017). No case for case in locality: Case does not help interpretation when intervention blocks A-bar chains. *Glossa: a Journal of General Linguistics*, *2*(1), 33. doi:10.5334/gjgl.165

Friedmann, N., & Shapiro, L. P. (2003). Agrammatic comprehension of simple active sentences with moved constituents: Hebrew OSV and OVS structures. *Journal of Speech Language and Hearing Research*, *46*(2), 288–297. doi:10.1044/1092-4388(2003/023

Fyndanis, V., Varlokosta, S., & Tsapkini, K. (2010). Exploring wh-questions in agrammatism: Evidence from Greek. *Journal of Neurolinguistics*, *23*(6), 644–662. doi:10.1016/j.jneuroling.2010.06.003

Garraffa, M., & Grillo, N. (2008). Canonicity effects as grammatical phenomena. *Journal of Neurolinguistics*, *21*(2), 177–197. doi:10.1016/j.jneuroling.2007.09.001

Grodzinsky, Y. (1995). Trace deletion, θ-roles, and cognitive strategies. *Brain and Language*, *51*(3), 469–497. doi:10.1006/brln.1995.1072

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, *23*(01), 1–21. doi:10.1017/S0140525X00002399

Hagiwara, H., & Caplan, D. (1990). Syntactic comprehension in Japanese aphasics: Effects of category and thematic role order. *Brain and Language*, *38*(1), 159–170. doi:10.1016/0093-934X(90)90107-R

Halliwell, J. F. (2004). *Syntactic movement and Korean aphasic comprehension* (*Ph.D.*), Michigan State University.

Hanne, S., Burchert, F., De Bleser, R., & Vasishth, S. (2015). Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction. *Journal of Neurolinguistics*, *34*, 83–111. doi:10.1016/j.jneuroling.2014.12.003

Hanne, S., Burchert, F., & Vasishth, S. (2016). On the nature of the subject–object asymmetry in wh-question comprehension in aphasia: Evidence from eye tracking. *Aphasiology*, *30*(4), 435–462. doi:10.1080/02687038.2015.1065469

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB–eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau 62*(1), 10–20. http://doi.org/10.1026/0033-3042/a000029

Hickok, G., & Avrutin, S. (1996). Comprehension of wh-questions in two Broca's aphasics. *Brain and Language*, *52*(2), 314–327. doi:10.1006/brln.1996.0014

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. doi:10.1198/106186006X133933

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). *Aachener aphasie test (AAT): Handanweisung*. Göttingen: Verlag für Psychologie Hogrefe.

Kljajevic, V., & Murasugi, K. (2010). The role of morphology in the comprehension of *wh* -dependencies in Croatian aphasic speakers. *Aphasiology*, *24*(11), 1354–1376. doi:10.1080/02687030903515347

Koopman, H., & Sportiche, D. (1991). The position of subjects. *Lingua. International Review of General Linguistics. Revue internationale De Linguistique Generale*, *85*(2-3), 211–258. https://doi.org/10.1016/0024-3841(91)90022-W

Kornfilt, J. (2003). Subject case in Turkish nominalized clauses. In U. Junghanns, & L. Szucsich (Eds.), *Syntactic structures and morphological information* (pp. 129–215). Berlin: Mouton de Gruyter.

MacWhinney, B., Osmán-Sági, J., & Slobin, D. I. (1991). Sentence comprehension in aphasia in two clear case-marking languages. *Brain and Language*, *41*(2), 234–249. doi:10.1016/0093-934X(91)90154-S

Maviş, İ, & Toğram, B. (2009). Afazi Dil Değerlendirme Testi (ADD) kullanım yönergesi. *Ankara: Detay Yayınları.*

Nerantzini, M., Varlokosta, S., Papadopoulou, D., & Bastiaanse, R. (2014). *Wh* -questions and relative clauses in Greek agrammatism: Evidence from comprehension and production. *Aphasiology*, *28*(4), 490–514. doi:10.1080/02687038.2013.870966

Neuhaus, E., & Penke, M. (2008). Production and comprehension of wh-questions in German Broca's aphasia. *Journal of Neurolinguistics*, *21*(2), 150–176. doi:10.1016/j.jneuroling.2007.05.001

Özsoy, A. S. (2009). Turkish as a (non)-wh-movement language. In E. A. Csató, G Ims, J. Parslow, F. Thiesen, & E. Türker (Eds.), *Turcological letters to bernt brendemoen* (pp. 221–232). Oslo, Norway: Novus Forlag.

Pesetsky, D. (1987). Wh-in-Situ: Movement and unselective binding. In E. J. Reuland, & A. G. B. ter Meulen (Eds.), *The representation of (in)definiteness* (pp. 98–129). Cambridge, MA: MIT Press.

Salis, C., & Edwards, S. (2008). Comprehension of wh-questions and declarative sentences in agrammatic aphasia: The set partition hypothesis. *Journal of Neurolinguistics*, *21*(5), 375–399. doi:10.1016/j.jneuroling.2007.11.001

Sezer, B., & Sezer, T. (2013). *TS Corpus: Herkes için Türkçe Derlem*. Paper presented at the Proceedings of the 27th National Linguistics Conference, Antalya. Retrieved from http://tscorpus.com/

Sheppard, S. M., Walenski, M., Love, T., & Shapiro, L. P. (2015). The auditory comprehension of *Wh* -questions in aphasia:

Support for the intervener hypothesis. *Journal of Speech Language and Hearing Research*, *58*(3), 781–797. doi:10.1044/2015_JSLHR-L-14-0099

Stavrakaki, S., & Kouvava, S. (2003). Functional categories in agrammatism: Evidence from Greek. *Brain and Language*, *86*(1), 129–141. doi:10.1016/S0093-934X(02)00541-2

Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on!. *The R Journal*, *1*(2), 14–17.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348. doi:10.1037/a0016973

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*(02), 135–178. doi:10.1017/S0954394512000129

Taylan, E. E. (1984). *The function of word order in turkish grammar*. Berkeley: University of California Press.

Thompson, C. K., & Choy, J. J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research*, *38*(3), 255–283. doi:10.1007/s10936-009-9105-7

Thompson, C. K., Tait, M. E., Ballard, K. J., & Fix, S. C. (1999). Agrammatic aphasic subjects' comprehension of subject and object ExtractedWhQuestions. *Brain and Language*, *67*(3), 169–187. doi:10.1006/brln.1999.2052

van der Meulen, I., Bastiaanse, R., & Rooryck, J. (2005). Wh-questions in agrammatism: A movement deficit? *Stem-, Spraak-en Taalpathologie*, *13*(1), 24–36.

Vikner, S. (1995). *Verb movement and expletive subjects in the Germanic languages*. Oxford: Oxford University Press.

Yarbay-Duman, T., Altınok, N., Özgirgin, N., & Bastiaanse, R. (2011). Sentence comprehension in Turkish Broca's aphasia: An integration problem. *Aphasiology*, *25*(8), 908–926. doi:10.1080/02687038.2010.550629

Appendices

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

	Lemma frequency	
Verb used	Turkish	German
to smack	1.392	22.105
to kick	1.052	42.523
to kiss	20.000	3.874
to pull	513.211	38.862
to push	29.314	9.497
to chase	6.366	2.535
to bite	4.450	1.835
to tickle	626	256
to poke	894	12
to stroke	4.175	1.395
to choke	15.982	526
to pinch	207	600

Appendix B. List of sentences used in the experiments

Object *which* questions

1. *Turkish test:* Adam/kadın hangi adamı/kadını tokatlıyor?
   German test: Welchen Mann/Welche Frau schlägt der Mann/die Frau?
   Eng. Trans.: Which man/woman is the man/woman smacking?

2. Adam/kadın hangi adamı/kadını tekmeliyor?
   Welchen Mann/Welche Frau tritt der Mann/die Frau?
   Which man/woman is the man/woman kicking?

3. Adam/kadın hangi adamı/kadını öpüyor?
   Welchen Mann/Welche Frau küsst der Mann/die Frau?
   Which man/woman is the man/woman kissing?

4. Adam/kadın hangi adamı/kadını çekiyor?
   Welchen Mann/Welche Frau zieht der Mann/die Frau?
   Which man/woman is the man/woman pulling?

5. Adam/kadın hangi adamı/kadını itiyor?
   Welchen Mann/Welche Frau stößt der Mann/die Frau?
   Which man/woman is the man/woman pushing?

6. Adam/kadın hangi adamı/kadını kovalıyor?
   Welchen Mann/Welche Frau jagt der Mann/die Frau?
   Which man/woman is the man/woman chasing?

7. Adam/kadın hangi adamı/kadını ısırıyor?
   Welchen Mann/Welche Frau beißt der Mann/die Frau?
   Which man/woman is the man/woman biting?

8. Adam/kadın hangi adamı/kadını gıdıklıyor?
   Welchen Mann/Welche Frau kitzelt der Mann/die Frau?
   Which man/woman is the man/woman tickling?

9. Adam/kadın hangi adamı/kadını dürtüyor?
   Welchen Mann/Welche Frau piekt der Mann/die Frau?
   Which man/woman is the man/woman poking?

10. Adam/kadın hangi adamı/kadını okşuyor?
    Welchen Mann/Welche Frau streichelt der Mann/die Frau?
    Which man/woman is the man/woman stroking?

11. Adam/kadın hangi adamı/kadını boğuyor?
    Welchen Mann/Welche Frau würdt der Mann/die Frau?
    Which man/woman is the man/woman choking?

12. Adam/kadın hangi adamı/kadını çimdikliyor?
    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 tokatlıyor?
    German test: Wen schlägt der Mann/die Frau?
    Eng. Trans.: Who is the man/woman smacking?

14. Adam/kadın kimi tekmeliyor?
    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?

18. Adam/kadın kimi kovalıyor?
Wen jagt der Mann/die Frau?
Who is the man/woman chasing?

19. Adam/kadın kimi ısırıyor?
Wen beißt der Mann/die Frau?
Who is the man/woman biting?

20. Adam/kadın kimi gıdıklıyor?
Wen kitzelt der Mann/die Frau?
Who is the man/woman tickling?

21. Adam/kadın kimi dürtüyor?
Wen piekt der Mann/die Frau?
Who is the man/woman poking?

22. Adam/kadın kimi okşuyor?
Wen streichelt der Mann/die Frau?
Who is the man/woman stroking?

23. Adam/kadın kimi boğuyor?
Wen würdt der Mann/die Frau?
Who is the man/woman choking?

24. Adam/kadın kimi çimdikliyor?
Wen kneift der Mann/die Frau?
Who is the man/woman pinching?

Subject *which* questions

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

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

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

33. Hangi adam/kadın adamı/kadını dürtüyor?
Welcher Mann/Welche Frau piekt den Mann/die Frau?
Which man/woman is poking the man/woman?

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

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

36. Hangi adam/kadın adamı/kadını çimdikliyor?
Welcher Mann/ Welche Frau kneift den Mann/die Frau?
Which man/woman is pinching the man/woman?

Subject *who* questions

37. *Turkish test:* Kim adamı/kadını tokatlıyor?
German test: Wer schlägt den Mann/die Frau?
Eng. Trans.: Who is smacking the man/woman?

38. Kim adamı/kadını tekmeliyor?
Wer tritt den Mann/die Frau?
Who is kicking the man/woman?

39. Kim adamı/kadını öpüyor?
Wer küsst den Mann/die Frau?
Who is kissing the man/woman?

40. Kim adamı/kadını çekiyor?
Wer zieht den Mann/die Frau?
Who is pulling the man/woman?

41. Kim adamı/kadını itiyor?
Wer stößt den Mann/die Frau?
Who is pushing the man/woman?

42. Kim adamı/kadını kovalıyor?
Wer jagt den Mann/die Frau?
Who is chasing the man/woman?

43. Kim adamı/kadını ısırıyor?
Wer beißt den Mann/die Frau?
Who is biting the man/woman?

44. Kim adamı/kadını gıdıklıyor?
Wer kitzelt den Mann/die Frau?
Who is tickling the man/woman?

45. Kim adamı/kadını dürtüyor?
Wer piekt den Mann/die Frau?
Who is poking the man/woman?

46. Kim adamı/kadını okşuyor?
Wer streichelt den Mann/die Frau?
Who is stroking the man/woman?

47. Kim adamı/kadını boğuyor?
Wer würdt den Mann/die Frau?
Who is choking the man/woman?

48. Kim adamı/kadını çimdikliyor?
Wer kneift den Mann/die Frau?
Who is pinching 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:

fit <- cforest (as.factor(accuracy) ~., data = accuracy, control = cforest_unbiased (mtry = 3 ntree = 2000))

- Assessment of conditional variable importance is computed with:

varimp(fit, conditional = TRUE)

- A conditional inference tree is plotted with:

Tree <- ctree(as.factor(accuracy) ~., data = accuracy, control = cforest_unbiased); plot(Tree)