# German children's processing of morphosyntactic cues in *wh*-questions

ATTY SCHOUWENAARS
*University of Groningen and University of Oldenburg*

PETRA HENDRIKS
*University of Groningen*

ESTHER RUIGENDIJK
*University of Oldenburg*

ADDRESS FOR CORRESPONDENCE
Atty Schouwenaars, University of Oldenburg, Institute of Dutch Studies, Ammerländer Heerstraße 114-118, 26111 Oldenburg Germany. E-mail: a.schouwenaars@rug.nl

ABSTRACT
Two experiments investigated the effects of case and verb agreement cues on the comprehension and production of *which*-questions in typically developing German children (aged 7–10) and adults. Our aims were to determine (a) whether they make use of morphosyntactic cues (case marking and verb agreement) for the comprehension of *which*-questions, (b) how these questions are processed, and (c) whether the presence and position of morphosyntactic cues available for the listener influence the speaker's production of *which*-questions. Performance on a picture selection task with eye tracking shows that children with low working memory make less use of morphosyntactic cues than children with high working memory and adults when interpreting object questions. Gaze data of both groups reveal garden-path effects and revisions for object and passive questions, which can be explained by a constraint-based account. Furthermore, children's difficulties with object questions are related to the type of disambiguation cue. In a question elicitation task with patient-initial items, children overall prefer production of passives, whereas adults' productions depend on the availability of disambiguation cues for the listener.

Keywords: eye tracking; incremental processing; language acquisition; morphosyntax; *wh*-questions; working memory

It is well known that children have problems with thematic-role assignment. Understanding who is doing what to whom is crucial for the correct interpretation of complex sentences such as *wh*-questions. Many studies report that object questions, in which the object precedes the subject, are difficult for children to comprehend (e.g., for English, O'Grady, 1997; for Italian, De Vincenzi, Arduino, Ciccarelli, & Job, 1999; for Hebrew, Friedmann, Belletti, & Rizzi, 2009; Friedmann & Novogrodsky, 2011; for Dutch, e.g., Metz, van Hout, & van der Lely 2010; Schouwenaars, van Hout, & Hendriks, 2014; for German, Biran & Ruigendijk, 2015; Roesch & Chondrogianni, 2015). These studies often report

offline accuracy scores, investigating the final interpretation of *wh*-questions. Online self-paced-reading studies with adults report longer reading times for object questions than for subject questions (Meng & Bader, 2000; Schlesewsky, Fanselow, Kliegl, & Krems, 2000). Longer reading times are interpreted as a reflection of a revision necessary for the correct interpretation. In the current study, we will investigate online processing of *wh*-questions using eye tracking to find out whether gaze patterns reflect such revisions, not only for adults but also for children. Furthermore, children's working memory is measured, as processing *wh*-questions may involve keeping in mind several possible interpretations or maintaining the dislocated object in memory for some time, both of which require sufficient working memory capacity (e.g., Fiebach, Schlesewsky, & Friederici, 2002). German is chosen as the language of investigation because of its different morphosyntactic cues, such as case and verb agreement.

German allows for variation in word order, which makes the following sentence structurally ambiguous:

(1) *Welche Schüler begrüßen die      Lehrer?*
      Which pupils   are      greeting the teachers?

A native speaker of German, when reading this sentence out of context, will likely interpret *pupils* as the subject and *teachers* as the object of the sentence. This interpretation is guided by a preference for canonical word order in which the subject precedes the object. Nevertheless, also the reversed interpretation is possible, namely, teachers greeting pupils. Whereas German declarative sentences often start with the subject, *wh*-questions usually start with the *wh*-phrase. Accordingly, when the *wh*-phrase functions as the subject, the subject precedes the object, resulting in a subject question. In contrast, when the *wh*-phrase is the object, the object precedes the subject, resulting in noncanonical word order in object questions.

How does a listener know whether *pupils* or *teachers* is the subject of the sentence? In English, the position of the verb differs between subject and object questions (see the English translations of [2]). This does not hold for German, where the order is always noun phrase–verb–noun phrase (NP–V–NP) and hence does not help the listener in establishing the subject and the object of the sentence. Often context, prosody, or semantic cues such as definiteness and animacy help the listener to correctly interpret *wh*-questions (especially in globally ambiguous sentences like [1], see, e.g., Bouma, 2008). Moreover, morphosyntactic cues can disambiguate subject and object questions. For example, in German, case on the *wh*-word or article can disambiguate this as the subject or the object and lead to a single possible interpretation (see [2]).

(2a) *Welcher    Schüler begrüßt den    Lehrer?*
      Which$_{NOM}$ pupil   greets   the$_{ACC}$ teacher?
      "Which pupil is greeting the teacher?"
(2b) *Welchen    Schüler begrüßt der    Lehrer?*
      Which$_{ACC}$ pupil   greets   the$_{NOM}$ teacher?
      "Which pupil is the teacher greeting?"

Singular masculine nouns in German have distinctive case marking that can indicate the subject and the object of a sentence. Nominative case in (2a) on the *wh*-phrase, *welcher* "which," marks the first NP as the subject. Accusative case on the article of the second NP, *den* "the," marks the second NP as the object. Therefore, (2a) is a subject question. Likewise, accusative case in (2b) on the *wh*-phrase, *welchen* "which," marks this NP as the object and nominative case on the article of the second NP, *der* "the," marks this NP as the subject. Therefore, (2b) is an object question.

Verb agreement can also disambiguate subject and object questions. If only one NP agrees in number with the verb, only that NP can be the subject. In (3a) only the first NP, *welche Schülerin* "which pupil," corresponds in number with the singular inflection on the verb, *begrüßt* "greets," and therefore is the subject. This leads to a subject question. In (3b), only the second NP, *die Lehrer* "the teachers," corresponds in number with the plural inflection on the verb, *begrüßen* "greet," and therefore is the subject. This leads to an object question.

(3a) *Welche Schülerin begrüßt   die Lehrer?*
     Which   pupil$_{SG}$   greets$_{SG}$   the teachers$_{PL}$?
     "Which pupil is greeting the teachers?"
(3b) *Welche Schülerin begrüßen die Lehrer?*
     Which   pupil$_{SG}$   greet$_{PL}$   the teachers$_{PL}$?
     "Which pupil are the teachers greeting?"

In (3), case does not disambiguate between subject and object, as the determiners of feminine and plural nouns have the same form for nominative and accusative case. These examples are therefore disambiguated by verb agreement only.

The meaning of sentence (3b) could also be realized as a passive question. In passives, unlike active sentences, not the thematic role of agent, but that of patient is realized as the subject. Hence the patient *Welche Schülerin* "which pupil" is the object in the active question (3b), but the subject in the passive question (4). Therefore, unlike the object questions (2b) and (3b), the passive question (4) starts with the subject.

(4) *Welche Schülerin wird       von den Lehrern gegrüßt?*
    Which   pupil$_{SG}$   is-being$_{SG}$   by   the teachers$_{PL}$ greeted$_{PPART.}$?
    "Which pupil is being greeted by the teachers?"

Subject-first structures are acquired earlier and are easier to process than object-first structures. This difference is generally referred to as the subject–object asymmetry. Passives are generally regarded to be acquired relatively late in comprehension (Borer & Wexler, 1987; Maratsos, Fox, Becker, & Chalkley, 1985). Nevertheless, in passive questions thematic role assignment may be easier than in object questions, as passive morphology (the verb *werden* "to be," the by-agent, and the past participle) may be more noticeable and reliable than case or verb agreement. One reason to include passive questions in our study is to compare two different types of noncanonicity: object-before-subject and patient-before-agent. In object questions, these syntactic functions and thematic roles go together (as the subject is the agent), but in passive questions they do not.

Passive questions are a viable alternative to object questions for expressing a question about the patient.

In addition, we examine the production of questions. Comprehension may affect production. When in production multiple forms express the same meaning, the speaker's choice may be influenced by the listener's ease of comprehension. If speakers take the listener's perspective into account, we expect them to produce the form that is easier to comprehend for the listener. The presence and position of morphosyntactic cues may therefore not only influence comprehension but also indirectly production.

Thus, the research questions we address in this study are (a) whether German children and adults make use of morphosyntactic cues (case marking and verb agreement) for the comprehension of *which*-questions, (b) how *which*-questions are processed, and (c) whether the presence and position of morphosyntactic cues available for the listener influence the speaker's production of *which*-questions. These questions will be investigated in a picture-selection task using eye tracking, and a corresponding question-elicitation task with the same participants. We will first review previous explanations for the subject–object asymmetry in children's comprehension of *which*-questions. Next, we will review a potential account of children's production of *which*-questions. Predictions of this constraint-based account will be formulated for the final interpretations, online gaze patterns, and produced forms by adults and children, and for active as well as passive questions. Then, we will describe our experiment to test these predictions and present our behavioral results, gaze data, and production results. Finally, we will discuss the results and draw conclusions.

## EXPLAINING CHILDREN'S SUBJECT–OBJECT ASYMMETRY IN COMPREHENSION

German-speaking children's ability to use case marking for sentence comprehension starts to develop around the age of 5 (e.g., Lindner, 2003; Roesch & Chondrogianni, 2015). Nevertheless, even older children still make many mistakes (Biran & Ruigendijk, 2015). Whereas children interpret subject questions correctly, they often incorrectly interpret object questions as subject questions. It is argued that 3-year-old children are sensitive to differences in case marking, but are not yet able to use this for building the correct underlying syntactic structure (Schipke, Knoll, Friederici, & Oberecker, 2012). Children seem to be even less able to use verb agreement, as they still misinterpret object questions disambiguated solely by verb agreement until the age of 8 or 9 (for Dutch, Metz et al., 2010; Schouwenaars et al., 2014; for Italian, De Vincenzi et al., 1999), even though 5-year-old children seem sensitive to verbal inflection (Brandt-Kobele & Höhle, 2014). Object-first sentences disambiguated by verb agreement also seem to cause greater processing difficulties for German-speaking children than sentences disambiguated by case marking (Arosio, Yatsushiro, Forgiarini, & Guasti, 2012). The same holds for adults (Friederici, Steinhauer, Mecklinger, & Meyer, 1998; Meng & Bader, 2000). It has been argued that this

is caused by the fact that case marking appears directly on the NPs, whereas agreement markers on the verb are indirect (Clahsen, 1986), meaning that for agreement, number marking on the NP and number marking on the verb have to be linked to one another.

Various explanations have been proposed for children's subject–object asymmetry in comprehension. One explanation is a *processing explanation* known as the active filler hypothesis (AFH; Frazier & Flores d'Arcais, 1989), which has been extended to acquisition (Avrutin, 2000; Deevy & Leonard, 2004). When parsing a sentence, children (like adults) take the first NP to be the subject, which is assigned the agent role. For subject questions, this is the correct interpretation, but for object questions, it is incorrect. Once this misinterpretation is noticed, the parser has to go back to the beginning of the sentence and reinterpret the sentence. It is argued that children do not have enough working memory resources or cognitive control to do so (Choi & Trueswell, 2010; Deevy & Leonard, 2004). This explanation accounts for adults' and children's difficulties in processing object-first structures. It also makes predictions for incremental interpretation: in both subject and object questions, initially the first NP will be interpreted as the subject and hence agent. In the literature on AFH, no explicit predictions have been formulated on the processing of passive questions or on the production of *wh*-questions.

Another prominent explanation is a *syntactic explanation* derived from Rizzi's (1990, 2004) relativized minimality approach (RM; Friedmann et al., 2009; Friedmann & Novogrodsky 2011; Jakubowicz, 2011). RM posits that *wh*-questions involve syntactic movement operations. In object questions, a relation or dependency needs to be formed between the sentence-initial object *wh*-phrase and its trace in its original position. This becomes harder if there is an intervener (here the subject) that is a potential candidate for this dependency. Therefore, object questions are harder to process than subject questions, in which there is no intervener. Children experience difficulties especially when the object *wh*-phrase and the subject intervener are of the same structural type: for example, when they both have a determiner (article, *wh*-word) and a noun (see Friedmann et al., 2009, for details). According to RM, in passives there is no intervener (Contemori & Belletti, 2014). Instead, the internal argument is first "smuggled" inside the moved verb phrase beyond the position of the external argument. Then the internal argument is extracted from the verb to a higher position. Thus, the internal argument is closest to the subject position without directly crossing over the external argument (see Collins, 2005, for details). Assuming this smuggling hypothesis, no interpretation problems are predicted for passive questions. Furthermore, it is argued that children prefer passive constructions over object-first constructions in production (Jensen de López, Sundahl Olsen, & Chondrogianni, 2014). Therefore, RM predicts difficulties in comprehension and production of object questions, but not subject and passive questions. RM as a theoretical account has been used to predict slower sentence processing for intervention effects, but it does not make predictions about the exact locus of processing difficulty.

A third prominent explanation for children's subject–object asymmetry is a *cue-based explanation* based on the competition model (CM). The CM posits that

Applied Psycholinguistics 39:6
Schouwenaars et al.: German children's processing of morphosyntactic cues in *wh*-questions

1284

people compute the interpretation of a sentence on the basis of various linguistic cues, eventually choosing the interpretation with the highest likelihood. Initially introduced for sentence processing, this performance model was later applied to language acquisition (Bates & MacWhinney, 1989; MacWhinney, 2005). According to the CM, language acquisition requires detecting surface cues in the language and determining the relative strength of these cues, which is based on the reliability and availability of the cues. Whereas there is consensus that case cues are more reliable than word order cues in German, there is no agreement on the validity (the product of reliability and availability) of these cues. According to Kempe and MacWhinney (1998), the validity for word order is higher than for case, whereas Dittmar, Abbot-Smith, Lieven, and Tomasello (2008) argue that the validity for case is higher than for word order in German. Regarding acquisition, the CM predicts that children acquire cues with a higher validity before those with a lower validity (Bates & MacWhinney, 1987). Furthermore, children's interpretations initially seem to depend on cue availability, and only later cue reliability is used. This could be an explanation for children's difficulties interpreting object questions. When they base their interpretation on cues that are high in availability, such as word order, instead of high in reliability, such as case, they interpret object questions as subject questions. The CM also makes explicit predictions about children's comprehension of passives. Due to language-specific properties, such as less reliance on constituent order in German than in English, it is predicted that German children understand passive sentences 1 year earlier than English children (Aschermann, Gülzow, & Wendt, 2004). To our knowledge, however, there are no studies within CM directly comparing comprehension of passives and object-first structures.

Regarding sentence processing, earlier work argues that interpretations do not change when new information comes in (MacWhinney, Bates, & Kliegl, 1984), but later work argues that thematic role assignment is updated at each point in sentence processing and therefore interpretations can change (Bates & MacWhinney, 1989). As for sentence production, according to the CM this is determined by function and frequency of grammatical forms (Bates & MacWhinney, 1989). Therefore, predictions about the production of *wh*-questions cannot directly be derived from the model itself.

Another explanation for children's subject–object asymmetry in comprehension is a *constraint-based explanation* in terms of optimality theory (OT; see Prince & Smolensky, 2004). In OT, the realization and interpretation of linguistic expressions is determined by the interaction between the constraints of the grammar, which express general tendencies of the language that can be in conflict. The realized form or selected interpretation is the form or interpretation that optimally satisfies these interacting constraints. Children's interpretation of *wh*-questions may result from the interaction between conflicting constraints (Schouwenaars et al., 2014). A first relevant constraint is WH-FIRST, which holds that a *wh*-constituent comes first in a sentence. When the *wh*-consituent is the patient, this constraint is in conflict with the constraint AGENT-FIRST, which holds that the agent comes first in a sentence (cf. Bouma, 2008; de Hoop & Lamers, 2006). Because, in German, WH-FIRST is ranked higher than AGENT-FIRST

(Bouma, 2008; Zeevat, 2006), a violation of the weaker constraint AGENT-FIRST is allowed in order to satisfy the stronger constraint WH-FIRST. Other morphosyntactic constraints outranking AGENT-FIRST are CASE, which holds that the subject is marked with nominative case and the object is marked with accusative case, and AGREEMENT, which holds that the verb agrees with the subject (de Hoop & Lamers, 2006). As a result of these interacting constraints, the optimal interpretation of object questions and passive questions, satisfying the constraints best, is a patient-first interpretation.

In OT, children are argued to initially entertain a different constraint ranking than adults (e.g., Fikkert & de Hoop, 2009; Smolensky, 1996). This explains children's non-adultlike patterns of production and interpretation. For example, unlike adults, children may give more importance to AGENT-FIRST than to AGREEMENT (Schouwenaars et al., 2014) and CASE. This non-adultlike ranking leads to a different optimal interpretation for object questions, namely, an agent-first interpretation.

OT is also able to make empirically testable predictions about the interpretation of incomplete sentences, and thus about incremental word-by-word processing (see de Hoop & Lamers, 2006; Stevenson & Smolensky, 2006). As some constraints only become relevant later in the sentence, when linguistic cues become available that allow potential outputs to be evaluated on the basis of these constraints, intermediate interpretations may differ from final interpretations.

## CHILDREN'S PRODUCTION OF *WH*-QUESTIONS

It is unclear whether the subject–object asymmetry found for comprehension also extends to production. For English, for example, Stromswold (1995) found that English children started producing object and subject questions at the same age (between age 1 year, 8 months [1;8] and 3;8) in spontaneous speech. Schouwenaars et al. (2014), in a *wh*-question elicitation task, found that Dutch 6- and 7-year-olds did not make mistakes in their production of object questions, although they, like adults, preferred to produce passive questions (70%). For Italian, no differences were found between 3- to 5-year-old children's productions of subject and object *which*-questions in a *wh*-question-elicitation task (Guasti, Branchini, & Arosio, 2012). Nevertheless, besides object questions (~30%), children produced alternative questions with clefts, putting the subject in dislocated position (~20%) or dropping the argument (~45%), which can be explained as a strategy to avoid object questions. This avoidance may indicate that children have problems producing object questions. In Hebrew, 3- and 4-year-old children avoided object relative clauses and produced more subject relatives in a relative-clause-elicitation task (Friedmann et al. 2009), which the authors argue is similar to the subject–object asymmetry in comprehension. Likewise, Italian children (as well as adults) produce passives instead of object relatives (Belletti & Contemori, 2010). No *wh*-question elicitation study has so far been reported with German children. Biran and Ruigendijk (2015) report that German children repeated fewer *which*-object questions correctly than *which*-subject questions on a repetition

task. Often children changed the object-first sentence into a subject-first sentence. Note, however, that the repetition method does not purely test production; to repeat a sentence, it must be understood to a certain degree as well.

As mentioned above, the AFH, RM, and the CM have been proposed to explain subject–object asymmetries in children's comprehension of *wh*-questions. However, these accounts do not make explicit predictions about children's production of *wh*-questions and require additional mechanisms to explain children's performance in production. OT, in contrast, makes explicit predictions about production as well. Constraints in OT can be applied to a set of potential meanings to select the optimal meaning for a given form (as in comprehension). However, these constraints can also be applied to a set of potential forms to select the optimal form for a given meaning (as in production). A general assumption in OT is that comprehension and production are explained by the same grammar (i.e., the same set of constraints under the same ranking). Although the constraints are the same, they can nevertheless have different effects in comprehension and production because they apply to different potential outputs (meanings and forms, respectively; see Hendriks, 2014; Smolensky, 1996). Thus, OT may make different predictions for the comprehension and production of *which*-questions (see Schouwenaars et al., 2014).

## PREDICTIONS FOR OUR STUDY

To examine how adults and children interpret, process, and produce *which*-questions, in an eye-tracking experiment we collect responses indicating final interpretations as well as gaze data revealing midsentence interpretations, and in a production experiment we elicit questions. In this section, predictions about the outcomes of the experiment are presented based on the constraint-based OT account, as this allows us to formulate specific predictions about adults' and children's final interpretations, their incremental processing, as well as their production of *wh*-questions. As some of the OT constraints reflect well-accepted views on *wh*-questions, these predictions are not necessarily incompatible with the other three models discussed above.

Regarding children's final interpretations of *which*-questions, if children incorrectly have ranked the AGENT-FIRST constraint highest and thus prefer the first NP to be the agent, this results in an adultlike agent-first interpretation of subject questions. Object questions, in contrast, are predicted to receive an incorrect agent-first interpretation. For passive questions, an incorrect agent-first interpretation is predicted too. Nevertheless, the interpretation of passive questions may be less affected than that of object questions, as passive questions contain multiple cues for interpretation (the verb *werden* "to be," the *by*-phrase, and the past participle). As these cues may be targeted by other constraints, not discussed here for reasons of space, children may base their interpretations on these other constraints and thus interpret passive questions correctly.

Turning to adults' incremental processing of *which*-questions, there are three important moments in the sentence. Consider an object question disambiguated by verb agreement such as (3b). First, when the singular *wh*-phrase *welche*

*Schülerin* "which pupil" is encountered, the agent-first interpretation is the optimal interpretation: it satisfies AGENT-FIRST, and CASE and AGREEMENT cannot be evaluated at this point, since there is no overt case marking and no verb yet. Then, when the plural verb *begrüßen* "greet" is encountered, the patient-first interpretation becomes the optimal interpretation: the agent-first interpretation now violates AGREEMENT because the sentence-initial *wh*-phrase and the finite verb do not agree in number, and although the patient-first interpretation violates AGENT-FIRST, this interpretation is nevertheless optimal because AGREEMENT is ranked higher than AGENT-FIRST. Finally, when the second NP, the plural *die Lehrer* "the teachers," is encountered, the patient-first interpretation remains optimal: this interpretation satisfies AGREEMENT because the finite verb agrees with the second NP. Thus, in object questions disambiguated by verb agreement, a shift is predicted from an agent-first interpretation at the sentence-initial *wh*-phrase to a patient-first interpretation at the finite verb.

Also in subject questions, the initial interpretation is guided by AGENT-FIRST. As the initial agent-first interpretation does not violate any further constraints, the intermediate and final interpretations are the same and no shift in interpretation is predicted.

For passive questions, the initial interpretation at the *which*-phrase is also determined by AGENT-FIRST, resulting in an agent-first interpretation. Next, the verb *wird* "is being" is encountered, indicating a passive question.[1] In passives, the patient is the subject, and therefore the verb must agree with the patient and not with the agent. Due to a violation of AGREEMENT by the agent-first interpretation, now the patient-first interpretation becomes the optimal interpretation and remains optimal. Therefore, in passive questions a shift is predicted from an agent-first interpretation to a patient-first interpretation at the finite verb.

The constraint-based OT account predicts a shift in interpretation midsentence for object questions disambiguated by verb agreement and for passive questions, but not for subject questions disambiguated by verb agreement. Further, the constraint-based OT account predicts no intermediate shifts in interpretation for object questions disambiguated by case on the first NP. Because CASE is ranked higher than AGENT-FIRST, the patient-first interpretation is the optimal interpretation already at the *wh*-phrase and remains optimal when encountering the next words. If children have ranked AGENT-FIRST too high compared to adults', they will show initial agent-first interpretations. Then, in contrast to adults, children may not overcome their initial misinterpretation of object questions and passive questions if neither AGREEMENT nor CASE outranks AGENT-FIRST.

As mentioned above, OT also makes specific predictions about the production of *which*-questions. When speakers wish to express a question about the agent (i.e., a question in which the *wh*-constituent is the agent), the optimal form is a subject question. This form satisfies all constraints mentioned, as the *wh*-constituent as well as the agent is in the first position. When speakers wish to express a question about the patient, there are two optimal forms: object questions and passive questions. Both forms violate the AGENT-FIRST constraint, but as subject questions violate higher ranked constraints, this form is suboptimal. Therefore, optionality is predicted: speakers can use two different forms to express the same meaning, namely, object questions and passive questions.

If the speaker takes into account the listener's perspective, the speaker is expected to choose the form that is easiest to understand for the listener. For example, the object question form *Welchen Schüler begrüßt der Lehrer?* "Which pupil-ACC is the teacher-NOM greeting?" starts with a masculine NP carrying accusative case. As case is unambiguously specified, leading to a correct object-initial interpretation already at the *wh*-phrase, no shift in interpretation occurs, and hence the sentence should be relatively easy to understand for listeners. In contrast, with feminine, neuter, or plural NPs case morphology does not unambiguously specify whether the first NP is subject or object, then a passive question is predicted to be easiest to understand, because passive questions contain more disambiguating cues in the form of passive morphology than object questions. We therefore predict that when case is available as an early disambiguation cue, speakers who take into account their listener will more likely produce an object question, whereas when case is not available, speakers will more likely produce a passive question. A key question is whether children as speakers are capable of taking into account the perspective of the listener.

Summarizing, the constraint-based OT account predicts the following: (a) adults initially incorrectly interpret, and subsequently revise their interpretation of, passive questions and object questions disambiguated by verb agreement, whereas no revisions are predicted for subject questions and for object questions disambiguated by case; (b) children incorrectly interpret object questions as subject questions; and (c) speakers produce subject questions when the *wh*-constituent is the agent, object questions when the *wh*-constituent is the patient and case marking is available, and passive questions when the *wh*-constituent is the patient and no case marking is available, as these latter forms are assumed to be easiest to understand for the listener.

## CURRENT STUDY

To examine these predictions, we conducted an eye-tracking experiment and a production experiment. To avoid an effect of syntactic priming, the experiments were carried out in two sessions with at least 3 days between them. In the first session comprehension was tested, and in the second session production. We will first present the comprehension experiment and then the production experiment.

## EXPERIMENT 1: COMPREHENSION

We investigate how German children and adults understand and process *which*-questions, and to what extent and when they make use of case and verb agreement cues in their interpretation of *which*-questions.

### Method

*Participants.* Thirty-six typically developing children with no diagnosed language, hearing, or speech pathologies (as reported by the parents) between the age of 7 and 10 were tested (22 male, 7;05–10;09, $M = 9;01$ years old, $SD = 12.7$ months). As a control group 30 adults were tested (14 male, $M = 24$

years old, $SD = 31.5$ months). Participants were recruited at and around the University of Oldenburg. They gave written informed consent prior to the experiment. The study was approved by the Ethical Committee of the University of Oldenburg and in accordance with the declaration of Helsinki.

### Screening tests.

AUDITORY DISCRIMINATION OF CASE.    In a first screening test, children's discrimination of nominative and accusative case marking on determiners was tested in an auditory discrimination test. Stimuli were presented auditorily and consisted of pairs of question words or determiners (as in [4]) and pairs of NPs (as in [5]), which were either the same (4) or different with respect to case (5).

(4) der-der
(5) welcher Hund-welchen Hund

The participants had to press a button marked with *gleich* (the same) or *nicht gleich* (not the same) depending on whether the two words or NPs were the same or not. In total 16 pairs were presented; 8 per condition (same vs. different). One of the 36 children did not pass this test on a criterion of 14 or more out of 16 correct ($M = 97.4$, $SD = 4.51$; 25 children made no mistakes, 8 children made one mistake, 2 children made two mistakes, and 1 child made three mistakes).

VERB AGREEMENT.    To ensure that children understood verb agreement in declarative sentences in which word order does not play a role, a second screening test involving a picture-selection task was carried out. A pair of pictures was presented on the screen while a prerecorded sentence was presented auditorily. The children were asked to select the picture that best matched the sentence (see [6] and Figure 1).

(6) *Sie        malt/malen        die Prinzessin*.
    pronoun$_{SG/PL}$  paint$_{SG}$/paint$_{PL}$  the princess
    "She/They paint(s) the princess."

The German pronoun *sie* is ambiguous and can refer to a singular feminine referent ("she") or a plural referent ("they"). In these sentences, therefore, the
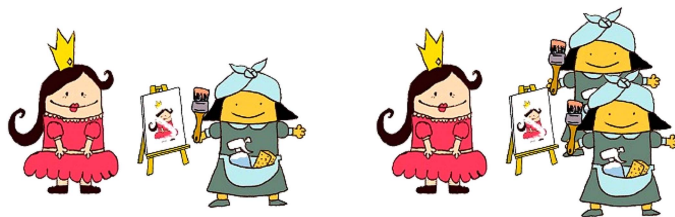


Figure 1. Example of a picture pair, one matching the single-subject interpretation (left), and the other matching the plural-subject interpretation (right) of sentence (6) *Sie malt/malen die Prinzessin* "She/they paint(s) the princess."

number of the subject referent is exclusively determined by the number marking on the finite verb. Each picture pair consisted of one picture corresponding to the singular interpretation of the subject and another corresponding to the plural interpretation of the subject (see Figure 1). The position of the target picture (left or right) and of the agent referent on the pictures was balanced over four lists. We used a total of 16 items; 8 per condition (singular vs. plural), with four reversible transitive verbs (*filmen* "to film," *fangen* "to catch," *malen* "to paint," and *waschen* "to wash"). The third-person singular form for the verbs *filmen* and *malen* are formed by stem + t, and for the verbs *fangen* and *waschen* by vowel-change in the stem + t. The latter may be more salient and therefore better distinguishable from the plural form. Both types of verbs were at ceiling level (no vowel-change: $M = 96.7\%$, $SD = 1.79$; vowel-change: $M = 96.2\%$, $SD = 1.91$). Only 1 of the 36 children did not pass this screening test on a criterion of scoring at least 14 out of 16 items correct (19 children made no mistakes, 12 children made one mistake, 4 children made two mistakes, and 1 child made three mistakes).

One child failed on the auditory discrimination of case screening test and another child on the verb agreement screening test. These children are excluded from further analysis. Of the remaining 34 children (21 male, 7;05–10;09, $M = 9;01$ years old, $SD = 12.7$ months) we can be sure that they perceive the differences in case morphology on determiners and *wh*-words and are sensitive to the number information provided by verbal inflection.

DIGIT SPAN TEST.   To examine the role of processing capacity in the comprehension of *wh*-questions, children's working memory was tested with a digit span test (HAWIK-IV; Petermann & Petermann, 2007) in two conditions: forward and backward. The child was asked to repeat a sequence of digits from 1 to 9, which was read out loud by the experimenter, in the given order (forward) or in the reversed order (backward). The forward session started with a sequence of three digits, the backward session with a sequence of two. For each sequence length, there were two trials, after which the number of digits in a sequence increased with one more digit. The test ended when both trials of the same length were recalled incorrectly. For the analyses, we used the backward digit span (number of digits of longest sequence recalled in reversed order correctly), because besides temporary storage (remembering the digits) it also requires manipulation of information (reordering the digits) and hence is considered a more complete measure of working memory (Baddeley, 2003).

*Comprehension of which-questions.*
STIMULI.  A picture selection task with eye tracking was used to test the comprehension of three different types of *which*-questions: subject *which*-questions, object *which*-questions, and passive *which*-questions (see [7]–[15] in Table 1). The subject and object questions were disambiguated by only case, only agreement, or both, resulting in six conditions in total. The differences between these conditions were realized by the gender and number of the nouns. Determiners of German singular masculine nouns differ between nominative (*der*) and

Table 1. *Example of test sentences*

| Type of question | Type of cue | | Examples of different conditions |
|---|---|---|---|
| Subject (SVO) | *Case* | (7) | *Welcher Esel wäscht den Pinguin?* <br> Which$_{NOM}$ donkey$_{SG}$ wash$_{SG}$ the$_{ACC}$ Penguin$_{SG}$ <br> "Which donkey is washing the penguin?" |
| | *Agr* | (8) | *Welche Giraffe wäscht die Kühe?* <br> Which$_{AMB}$ giraffe$_{SG}$ wash$_{SG}$ the$_{AMB}$ cow$_{PL}$ <br> "Which giraffe is washing the cows?" |
| | *AgrCa* | (9) | *Welche Füchse waschen den Schwan?* <br> Which$_{AMB}$ fox$_{PL}$ wash$_{PL}$ the$_{ACC}$ swan$_{SG}$ <br> "Which foxes are washing the swan?" |
| Object (OVS) | *Case* | (10) | *Welchen Esel wäscht der Pinguin?* <br> Which$_{ACC}$ donkey$_{SG}$ wash$_{SG}$ the$_{NOM}$ Penguin$_{SG}$ <br> "Which donkey is the penguin washing?" |
| | *Agr* | (11) | *Welche Giraffe waschen die Kühe?* <br> Which$_{AMB}$ giraffe$_{SG}$ wash$_{PL}$ the$_{AMB}$ cow$_{PL}$ <br> "Which giraffe are the cows washing?" |
| | *AgrCa* | (12) | *Welche Füchse wäscht der Schwan?* <br> Which$_{AMB}$ fox$_{PL}$ wash$_{SG}$ the$_{NOM}$ swan$_{SG}$ <br> "Which foxes is the swan washing?" |
| Passive | Passive morphology *Pas(a)* | (13) | *Welcher Esel wird von dem Pinguin* <br> Which$_{NOM}$ donkey$_{SG}$ is-being$_{SG}$ by the$_{DAT}$ penguin$_{SG}$ <br> *gewaschen?* <br> wash$_{PPART}$ <br> "Which donkey is being washed by the penguin?" |
| | Passive morphology *Pas(b)* | (14) | *Welche Giraffe wird von den Kühen* <br> Which$_{AMB}$ giraffe$_{SG}$ is-being$_{SG}$ by the$_{DAT}$ cow-PL <br> *gewaschen?* <br> wash$_{PPART}$ <br> "Which giraffe is being washed by the cows?" |
| | Passive morphology *Pas(c)* | (15) | *Welche Füchse werden von dem Schwan* <br> Which$_{AMB}$ fox$_{PL}$ are-being$_{PL}$ by the$_{DAT}$ swan$_{SG}$ <br> *gewaschen?* <br> wash$_{PPART}$ <br> "Which foxes are being washed by the swan?" |

accusative case (*den*), while no such distinction is present in determiners for feminine or plural nouns (for both cases *die*). In the first condition Case, masculine noun pairs provided the case disambiguation cue on both the initial *wh*-phrase and the second NP. Both nouns were singular, so verb agreement was not available as a cue (see [7] and [10]). In a second condition *Agr*, feminine noun pairs were used, so case was not available as cue. The first noun pair was singular and the second noun pair was plural to provide the subject–verb agreement disambiguation cue (see [8] and [11]). To examine whether a case disambiguation cue in addition to an agreement disambiguation cue helps the listener to revise a

first interpretation, a third condition was tested. In this condition, AgrCa, questions were disambiguated by subject–verb agreement and case on the second NP. Of these noun pairs, the first noun was masculine plural (thus ambiguously case marked) and the second noun was masculine singular, thus providing the subject–verb disambiguation cue and a case marking cue on the second NP (see [9] and [12]). With respect to the timing of the disambiguation cues, the *Case* condition has an early disambiguation cue on the first NP, whereas in the other two conditions, Agr and AgrCa, disambiguation takes place later in the sentence (see [10] vs. [11] and [12]).

For passive questions the same noun pairs were used as for active questions. In Pas(a) the first and the second NP are both masculine singular (see [13]). In Pas (b) the first NP is feminine singular and the second NP is feminine plural (see [14]). In Pas(c) the first NP is masculine plural and the second NP is masculine singular (see [15]). Nevertheless, for passive sentences these different nouns do not lead to a distinction with respect to type of disambiguation cue, as in active sentences. The passive questions were always disambiguated by passive morphology instead.

There were four lists that differed in order of the items and in position of the target picture (left or right). In total 54 test items were presented: 6 for every condition in Table 1, leading to 18 items per question type. For each trial two pictures were presented side by side. The pictures depicted the correct interpretation or the incorrect interpretation resulting from a role reversal. For example the left-sided picture in Figure 2 represents the correct patient-first interpretation of sentence (12). In the right-sided picture the thematic roles are reversed, representing the incorrect agent-first interpretation.

*Procedure.* In the familiarization phase, the participants were presented with a picture pair for 2500 ms to get used to the pictures. Next, a fixation cross appeared on the screen. After fixating the cross for 500 ms, the picture pair reappeared on the screen, and 50 ms later the prerecorded sentence was presented auditorily, after which the participants had to press the button corresponding to the picture they thought best fitted the sentence (see Appendix A for task
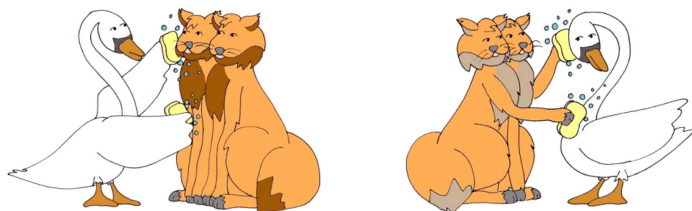


Figure 2. Example of a picture pair, with one picture matching the patient-first interpretation (left) and the other picture matching the agent-first interpretation (right) of sentence (12): *Welche Füchse wäscht der Schwan* "Which foxes is the swan washing?" Depending on the nouns used in the test sentences, the number of animals in the picture differs between two (one of each kind, in the Case condition) and three (one of one kind and two of the other kind in the Agr and AgrCa conditions; see this example).

instructions). There was no response time limit. The test items were divided into two blocks of 27 items each, both preceded by a 9-point calibration in Tobii and by two practice items (e.g., "Which bird is building a nest?"). Furthermore, in total 7 filler items with one animate noun (e.g., "Which kangaroo is shooting the ball?") were included. Between the blocks, the verb agreement screening test described above was carried out. The digit span task and the case screening test were carried out in the second session, respectively before and after the first block of the production task. Both sessions took around 30–45 min.

The participants sat in front of a 23-inch Tobii TX300 eye tracker with a resolution of 1920 × 1080 pixels and a screen response time of 5 ms. The eye tracker was connected to two computers. One computer ran the experiment with the software E-Prime 2.0 (Psychological Software Tools, Inc.) and collected the behavioral data. With the use of TET-calls in E-Prime the participants' eye movements at a sample rate of 300 Hz were collected from the second computer.

### Analysis

#### Accuracy data.
GENERALIZED LINEAR MIXED-EFFECTS REGRESSION MODELING (GLMER).   We used GLMER with the software R (version 3.1.2) to analyze the accuracy data. As a model building strategy, we choose parsimonious mixed models (Bates, Kliegl, Vasishth, & Baayen, 2015), as these models are more suitable for the typical sample sizes of psycholinguistic research (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Our accuracy models include a binomial dependent variable with a logit link function of *Item accuracy* and random intercepts for *Participant* and *Item*. The necessity of taking into account random slopes was assessed. The inclusion of factors was assessed by comparing the Akaike information criterion scores (Akaike, 1974). A decrease of at least 2 in the Akaike information criterion scores means that the inclusion of a factor significantly improves the goodness of fit of the model. Of the fixed factors, the first level is taken as the reference and each other level is contrasted with this baseline level. The order of the levels and the coding for group is children (baseline level) coded as –1 and adults as 1; for type of question the order is subject (baseline, –1), object (0), and passive (1); for position of target, left (baseline, –1) and right (1); and for type of cue, AgrCa (baseline level, –1), Agr (0), and Case (1). To compare the second with the third level and so on, multiple comparisons were made with the use of the glht function of the "mult-comp" package (Hothorn, Bretz, & Westfall, 2008), which corrects for multiple comparisons and gives adjusted *p* values.

#### Gaze data.
PREPROCESSING OF THE GAZE DATA.   Validity of the gaze data was rated by the eye tracker with a value of 0 or 1, meaning that the system is certain that all relevant data for both eyes or highly probable estimations for one eye were recorded. Only valid data points were included. No participants or trials had to be removed due to insufficient (<75%) valid data points. No selection was made based on offline accuracy of the trials. Instead, gaze data from both correct and

incorrect trials were included to present a more complete picture of how cues are processed in general. Gaze data was limited to 3000 ms after the onset of the stimulus to cover the complete range of time from onset until the average response time. Areas of interest (AOIs) were defined over *target* interpretation (target picture), *competitor* interpretation (competitor picture), and *not on AOI*. For the statistical analysis, the sum of looks to a specific AOI was calculated per participant per trial and per time bin of 200 ms from the raw data file. For the gaze plots, time bins of 50 ms were used for a more detailed picture.

GENERALIZED ADDITIVE MIXED MODELING (GAMM).   The gaze data were analyzed in R with GAMM (Wood, 2006, 2011) using the package mgcv 1.8.4 (Wood, 2006) and the package itsadug (van Rij, Baayen, Wieling, & van Rijn, 2015). GAMM is a nonlinear regression analysis and therefore particularly useful for time course data such as eye tracking (Nixon, van Rij, Mok, Baayen, & Chen, 2016; van Rij, Hollebrandse, & Hendriks, 2016). Like generalized linear mixed-effects regression modeling, GAMM allows for inclusion of both fixed and random factors. The crucial difference is that GAMMs manage nonlinear data sets. The relations between the factors and the dependent variable are modeled as smooth functions.[2] Smooth functions and parameters are determined by estimation procedures in order to avert overfitting and overgeneralization of the data (van Rij et al., 2016; Wood, 2006). For the model predictions we used difference plots from the itsadug package (van Rij et al., 2015). For example, the function *get_differences* and difference plots were used to calculate differences between children's and adults' looking behavior for subject, object, and passive questions.

### Results

We will present both offline accuracy scores and online gaze data. The offline accuracy scores inform us about the final interpretation the participants give to the *which*-questions. The online gaze data inform us about the processing during sentence presentation, namely, about the interpretations given to *which*-questions at different moments in time.

*Accuracy.*   Figure 3 shows the percentage of correct interpretations of *which*-questions for children (left) and adults (right). A GLMER model was made to compare the groups (children and adults). One by one, the following fixed factors were included to see whether they improved the goodness of fit of the model: group (adults vs. children), type of question (subject vs. object vs. passive), and type of cue (Case vs. Agr vs. AgrCa). The inclusion of type of cue (valid factor for subject and object questions only) did not improve the model. In addition, no interactions for this variable with group or type of question were found. We examined the possible effects of the material-related variables, such as verb, pair of nouns, session, direction of action, and position of target. Of these variables, only position of target (left vs. right) significantly improved the model. As position of target was balanced over type of question and type of cue, and changed for each item over the different lists, no interactions were found.
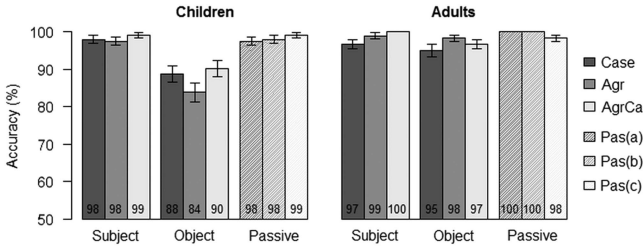
Figure 3. Percentages of correct interpretations of subject questions, object questions, and passive questions with their different cues. Case means disambiguated by case on the wh-phrase and the second NP, Agr means disambiguated by verbal agreement, AgrCa means disambiguated by verbal agreement and by case on the second NP, and Pas means passive construction. Error bars indicate standard error.

Table 2. *Fixed effects of best fitting generalized mixed effects model to fit the accuracy scores of the which-questions*

Formula

```
glmer(Item.ACC ~ Group * TypeOfQuestion + TargetPosition +
(1|Subject) + (1|Item), data = data, family = 'binomial')
```

| Predictor | Estimate | *SE* | *z* value | *p* value |
|---|---|---|---|---|
| (intercept) | 4.9176 | 0.4819 | 10.205 | < .001 |
| GroupAdults | –0.0080 | 0.6273 | –0.013 | .98 |
| Type of question obj | –2.3950 | 0.4026 | –5.949 | < .001 |
| Type of question pas | –0.0373 | 0.4821 | –0.078 | .94 |
| Position of target right | 0.4308 | 0.2065 | 2.086 | < .05 |
| GroupAdults*Type of Question obj | 1.4635 | 0.5616 | 2.606 | < .01 |
| GroupAdults*Type of Question pas | 1.0100 | 0.8182 | 1.234 | .22 |

Table 2 shows the final model for the overall analyses. With this model we can further investigate the effects of type of question and its interaction with group, which contain more than two levels and therefore require multiple comparisons. The only factor with two levels is the position of target. As shown in Table 2, items with the target picture on the right are interpreted better than those where it is on the left.

A multiple comparison reveals that there is a significant difference in accuracy between object questions and subject questions and between object questions and passive questions, but not between subject questions and passive questions as can be seen in Table 3.

The multiple comparison in Table 4 shows that the difference between the groups only holds for object questions and not for subject questions or passive questions: children score significantly worse than adults on object questions

Table 3. *Multiple comparisons of means for accuracy of the interpretation of the three different types of which-questions (Tukey contrasts)*

|  | Estimate | Standard error | $z$ value | $p$ value |
|---|---|---|---|---|
| Object-subject | –2.3951 | 0.4026 | –5.949 | < .001 |
| Passive-subject | –0.0374 | 0.4821 | –0.078 | .99 |
| Passive-object | 2.3577 | 0.3944 | 5.978 | < .001 |

Table 4. *Multiple comparisons of means for accuracy of the interaction between three different types of which-questions and group (Tukey contrasts)*

|  | Estimate | Standard error | $z$ value | $p$ value |
|---|---|---|---|---|
| Adults *Sub*-Children *Sub* | –0.0080 | 0.6271 | –0.013 | 1.000 |
| Children *Obj*-Children *Sub* | –2.3950 | 0.4024 | –5.951 | < .001 |
| Children *Pas*-Children *Sub* | –0.0374 | 0.4821 | –0.078 | 1.000 |
| Adults *Obj*-Adults *Sub* | –0.9315 | 0.4832 | –1.928 | .3598 |
| Adults *Pas*-Adults *Sub* | 0.9726 | 0.7132 | 1.364 | .7266 |
| Adults *Obj*-Children *Obj* | 1.4555 | 0.4835 | 3.010 | < .05 |
| Children *Pas*-Children *Obj* | 2.3575 | 0.3944 | 5.978 | < .001 |
| Adults *Pas*-Adults *Obj* | 1.9041 | 0.6608 | 2.881 | < .05 |
| Adults *Pas*-Children *Pas* | 1.0020 | 0.7766 | 1.290 | .7707 |

($\beta = 1.45$, $z = 5.978$, $p < .05$). Specifically, only for children there is a significant difference between subject and object questions ($\beta = -2.39$, $z = -5.951$, $p < .001$), but not for adults ($\beta = -0.93$, $z = -1.928$, $p = .359$). In contrast, the difference between object questions and passive questions is significant for both children ($\beta = 2.35$, $z = 5.978$, $p < .001$) and adults ($\beta = 1.90$, $z = 2.881$, $p < .05$).

A closer examination of children's accuracy scores for object questions revealed that most children (23 out of 34) made only one or no errors (out of 18 object question items). Four children scored at chance level or below when the object question was disambiguated by verb agreement only, but made only one or no errors when case or both case and agreement cues were available. Another 4 children scored at chance level or below for all types of cues. Three other children made two to six errors spread over all cue conditions.

To further unravel children's accuracy scores for object questions, we investigated the influence of two more factors: digit span backward and age. There was no correlation between digit span and age, $r$ (32) = .06, $p = .73$. Raw backward digit span was used to make three groups: *low* (digit span of 3, $n = 11$, 7-year-olds $n = 1$, 8-year-olds $n = 4$, 9-year-olds $n = 3$, and 10-year-olds $n = 3$), *medium* (digit span of 4, $n = 13$, 7-year-olds $n = 5$, 8-year-olds $n = 1$, 9-year-olds $n = 4$, and 10-year-olds $n = 3$), and *high* (digit span of 5–6, $n = 10$, 7-year-olds $n = 1$, 8-year-olds $n = 4$, 9-year-olds $n = 2$, and 10-year-olds $n = 3$). Groups
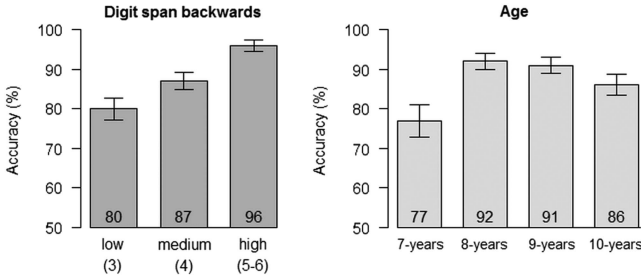
Figure 4. Children's mean accuracy scores (in percentages) on object questions per digit span group (left) and per age group (right).

Table 5. *Fixed effects of best fitting generalized mixed effects model to fit the accuracy scores of children's object questions*

**Formula**

```
glmer(Item.ACC ~ DSGroup + AgeGroups + (1|Codes),
data = dataSel, family = 'binomial', control = glmerControl
(optimizer = "bobyqa"))
```

| Predictor | Estimate | SE | z value | p value |
|---|---|---|---|---|
| (intercept) | 0.380 | 0.3173 | 1.199 | .2307 |
| Digit span medium | 0.874 | 0.2971 | 2.941 | < .01 |
| Digit span high | 1.874 | 0.4300 | 4.359 | < .001 |
| Age 8 | 1.433 | 0.4124 | 3.476 | < .001 |
| Age 9 | 1.298 | 0.3621 | 3.585 | < .001 |
| Age 10 | 0.758 | 0.3482 | 2.177 | < .05 |

instead of the scores as a range were used to avoid that correlations strongly depended on extreme values (as in this data only one child had a digit span score of 6). In order to see whether there were differences between different ages, age is divided into four groups: 7-year-olds ($n = 6$), 8-year-olds ($n = 9$), 9-year-olds ($n = 10$), and 10-year-olds ($n = 9$). Figure 4 shows the mean accuracy scores of object questions by children per digit span group (left) and per age group (right).

A new model is made with children's accuracy scores on object questions as a dependent variable. Because there was no correlation between children's age and their digit span scores, both digit span and age were included as fixed factors in the model. Only item was included as a random factor and not participant, because each participant had a single score of digit span and of age.

Table 5 shows the final model for the analysis of children's scores on object questions. Like the other models, this model contains variables with more than two levels. Therefore, multiple comparisons are made for the factors digit span (see Table 6) and age (see Table 7).

Table 6. *Multiple comparisons of means for children's accuracy scores on object questions in the three different digit span groups (Tukey contrasts)*

|  | Estimate | Standard error | *z* value | *p* value |
|---|---|---|---|---|
| Medium-low | 0.874 | 0.2971 | 2.941 | < .01 |
| High-low | 1.874 | 0.4300 | 4.359 | < .001 |
| High-medium | 1.000 | 0.4490 | 2.228 | .0643 |

Table 7. *Multiple comparisons of means for children's accuracy scores on object questions in the four different age groups (Tukey contrasts)*

|  | Estimate | Standard error | *z* value | *p* value |
|---|---|---|---|---|
| 8-7 | 1.433 | 0.4124 | 3.476 | < .01 |
| 9-7 | 1.298 | 0.3621 | 3.585 | < .01 |
| 10-7 | 0.758 | 0.3482 | 2.177 | .1287 |
| 9-8 | −0.135 | 0.4084 | −0.331 | .9875 |
| 10-8 | −0.675 | 0.3833 | −1.761 | .2908 |
| 10-9 | −0.540 | 0.3573 | −1.512 | .4288 |

The multiple comparisons in Table 6 confirm significant differences between the low digit span group and the two other groups. Children with a low digit span made more errors on the comprehension of object questions than children with a medium digit span ($\beta = 0.87$, $z = 2.941$, $p < .01$) and children with a high digit span ($\beta = 1.87$, $z = -4.359$, $p < .001$). Between the group of children with a medium and a high digit span no significant differences were found ($\beta = 1.00$, $z = -2.228$, $p = .0643$).

The multiple comparisons in Table 7 confirm that 7-year-old children made significantly more errors on the comprehension of object questions than 8-year-old children ($\beta = 1.43$, $z = 3.476$, $p < .01$) and 9-year-old children ($\beta = 1.29$, $z = 3.585$, $p < .01$). The difference between the 7-year-old and 10-year-old children was not significant ($\beta = 0.75$, $z = 2.177$, $p = .1287$). Also between the 8-, 9-, and 10-year-olds no significant differences were found.

Summarizing, the offline data show that children made significantly more errors than adults in their comprehension of object questions, but not of subject or passive questions. Children's comprehension of object questions was affected by digit span (children with a low digit span misinterpreted object questions significantly more often than children with a medium or high digit span) and age (7-year-olds misinterpreted object questions significantly more often than 8- and 9-year-olds). No differences were found with respect to the different disambiguation cues.

*Gaze data.* Sentence interpretation is an incremental process, which means that interpretation need not wait until the end of the sentence but can already take

place while words are encountered one by one. Crucially, the optimal inter-pretation can change over time. This is exactly what we will see in the gaze patterns for object and passive questions.

The gaze plots in Figure 5 show that for subject questions, children and adults look increasingly toward the target picture. For object questions, we first see an increase of looks toward the competitor picture, followed by an increase of looks toward the target picture. The increase of looks toward the target picture seems to be earlier for adults than for children. A similar pattern appears for passive questions.

A GAMM model is made to investigate differences between the two groups. In a later analysis, we will look at differences with respect to type of cue.

For our overall model we used TCDiff (the sum of looks toward the target minus the sum of looks toward the competitor picture) for timebins of 200 ms as the dependent variable. All interactions between group (adults vs. children) and type of question (subject vs. object vs. passive) were combined into one predictor to see whether there were differences between the groups with respect to the different types of questions. As random effect factors increase the time of running a model (which was already 12 hr), item was not included as a random effect factor. Instead, participant and type of question were combined into one random effect factor (ParticipantQuestion) and added to the model. A summary of the model is given in Appendix B (Table B.1). As this summary merely indicates whether the smooth of each variable is linear or not, further calculations are made in the following paragraphs.

The difference plots (see Figure B.1 in Appendix B) reveal differences between adults' and children's gaze patterns for object and passive questions, but not for subject questions. Children's looks toward the correct picture increase later than adults' for object and passive questions. The differences between children and adults lasted longer for object questions than for passive questions. This indicates that children needed more time than adults to revise the incorrect interpretation, and even more so in object questions than in passive questions.

To see whether different disambiguation cues lead to different gaze patterns for children and adults, we ran a second analysis. We visualized the gaze patterns for the object questions per type of cue for children and for adults (see Figure 6). For both children and adults, we clearly see a preference for the incorrect initial interpretation (more looks toward the competitor picture than toward the target picture) for the AgrCa and Agr conditions, but not for the Case condition.

To analyze whether these observed differences between the cues are significant, we made a second GAMM model. Now we included solely the data of the object questions. The input was again TCDiff for timebins of 200 ms. All interactions between group (adults vs. children) and type of cue (Case vs. Agr vs. AgrCa) were combined into one predictor. Participant was used as a random effect factor. A summary of the model is given in Appendix C (Table C.1).

Again difference plots were made to see whether the observed differences were significant (see Figures C.1 and C.2 in Appendix C). For children, there were significant differences in looks between object questions disambiguated by Case and the other two conditions (AgrCa and Agr). This is shown by the increasing
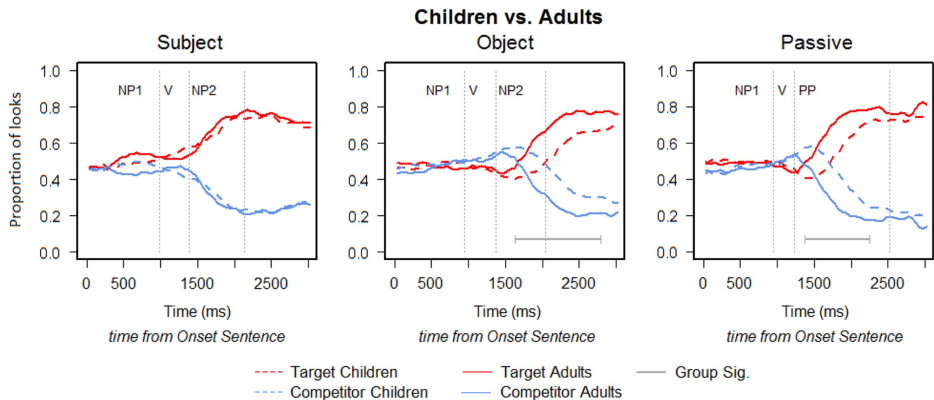
Figure 5. Children's (dashed line) and adults' (solid line) online gaze behavior for subject, object, and passive questions. The plots show separate lines for looks toward the target picture (red lines) and competitor picture (blue lines), for children (dashed lines) and adults (solid lines). The vertical lines indicate the mean onset of the verb, the mean onset of the second NP, and the mean offset of the sentence. The horizontal gray lines indicate a significant difference between children's and adults' gaze patterns analyzed with the statistical model described in the GAMM section.
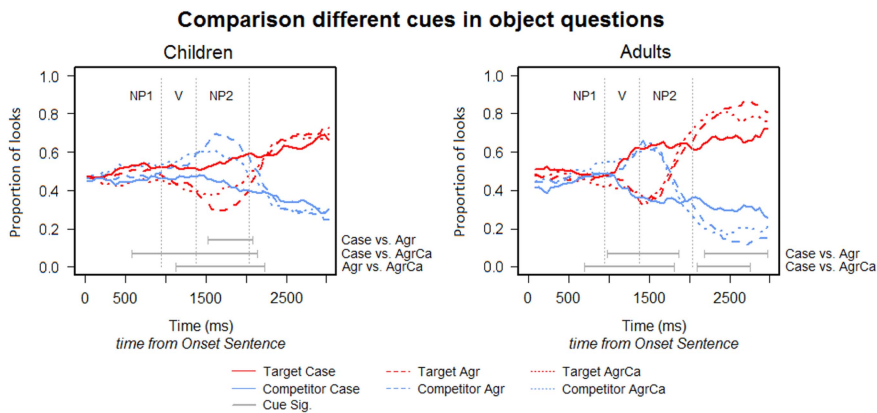
**Comparison different cues in object questions**



Figure 6. Children's (left plot) and adults' (right plot) online gaze behavior for object questions. The plots show separate lines for looks toward the target picture (red lines) and competitor picture (blue lines) per type of cue: AgrCa (dotted lines), Agr (dashed lines), and Case (solid lines). The vertical lines indicate the onset of the verb, the onset of the second NP, and the offset of the sentence. The gray horizontal lines indicate a significant difference between the types of cues.

proportion of looks toward the target picture for Case, whereas for AgrCa and Agr children initially showed an increasing proportion of looks toward the competitor picture, followed by an increasing proportion of looks toward the target picture. The same pattern and differences were found for adults. We take this to be an indication that case, in contrast to agreement, is used early in processing of *which*-questions. For children, an additional difference was found between the AgrCa and Agr conditions: the proportions of looks toward the competitor picture for object questions in the AgrCa condition was lower and dropped earlier than for the Agr condition. Thus, children, but not adults, seem to benefit from the extra case cue on the second NP.

Summarizing, the online gaze data show that both children's and adults' interpretation changes from an agent-initial interpretation to a patient-initial interpretation during the processing of object questions and passive questions. Children were slower in revising their initial interpretation than adults. Furthermore, whereas object questions disambiguated by verb agreement, or by verb agreement and case on the second NP, were initially interpreted as subject questions, object questions disambiguated by case on the first NP were not. These differences with respect to disambiguation cue may have implications for production when ease of comprehension is taken into account.

## EXPERIMENT 2: PRODUCTION

### Method

*Participants.*   Participants were the same as in the comprehension experiment.

*Materials and design.*   To test what type of questions children and adults produce, we conducted a question elicitation task that was modeled after the Diagnostic Evaluation of Language Variation test (Seymour, Roeper, & de Villiers, 2003). Every item includes a sequence of three pictures (Figure 7).

The first picture, together with the introductory sentence, presents the characters of the event. Two different characters of the same kind of animal are introduced to justify a *which*-question. The characters have different colors and are referred to as such. In the second picture, the action is shown. Here the crucial parts, either the agent(s) or the patient(s), are covered. This way, the participant can see which type of character is involved in the action, but not which of the introduced characters it is. Therefore, the participant has to ask a question, which is elicited with the accompanying sentence. This question has to start with a *which*-phrase (see Appendix A for the instruction). After the participant formulates the question, the answer is shown in the third picture.

The materials consist of 24 test items, preceded by 5 practice items. The practice items contain intransitive verbs with singular or plural agents. In the test items, the same types of questions are targeted as in the comprehension test. In half of the items, the agent of the picture is covered and in the other half, the patient. The same noun pairs are used as in the comprehension test, in order to see whether the use of case and/or agreement cues (8 items per cue) makes a
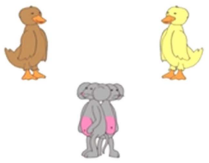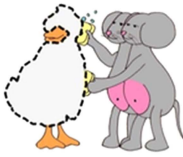
Applied Psycholinguistics 39:6
Schouwenaars et al.: German children's processing of morphosyntactic cues in *wh*-questions

1303

| Introduction | Question elicitation | Assertion elicitation |
|---|---|---|
| *Hier siehst du eine braune Ente, eine gelbe Ente und zwei Mäuse.* 'Here you see a brown duck, a yellow duck and two mice' | *Hier geht es ums waschen. Stell mir die richtige Frage und ich zeige dir die Antwort.* 'Here you see washing. Ask me the right question and I'll show you the answer.' | *Was passiert hier?* 'What happens?' |

Figure 7. Sample item for patient-initial questions in the elicitation task. A targetlike response could, for example, be an object question *Welche Ente waschen die Mäuse?* "Which duck are the mice washing?" or a passive question *Welche Ente wird von den Mäusen gewaschen?* "Which duck is being washed by the mice?"

difference in the participants' choice between a subject question, object question, or passive question.

*Procedure and scoring.* Participants' produced questions were recorded via E-prime and transcribed by two native German speakers. The targeted conditions were first divided into agent-initial or patient-initial, and subsequently divided into the different cue categories determined by the noun pairs (Case, Agr, and AgrCa). For every category, we scored participants' responses as *subject* questions, *object* questions, *passive* questions, *case errors* (involving incorrect or reversed case), *agreement errors* (involving incorrect number of the verb or NP), and *other* (other verbs used to describe the action, nonexistent noun, or verb forms, in situ questions).

### Results

For agent-initial items both children and adults produce subject questions. For patient-initial items adults produce roughly as many object questions (51%) as passive questions (48%). Children produce more passive questions (80%) and fewer object questions (13%). On the remaining items they make case errors, agreement errors, or produce other constructions. The case errors produced by children occur in both passive and object questions. A common pattern observed in passive questions is accusative instead of dative case marking in the *von*-NP construction (*Welche Gänse werden von den Fuchs getragen?* with incorrect *den* instead of correct *dem*). A common error pattern observed for patient-initial items is role reversal in which the first NP (the patient) has nominative case and the second NP (the agent) has accusative case.

   As we were interested in the production of object questions versus passive questions with respect to different disambiguation cues, we divide the patient-initial items according to the different cue conditions (see Figure 8).

   Children produced most object questions in the Case condition (24%). For the Agr and AgrCa conditions the percentages of produced object questions were lower (11% and 13%, respectively). Adults also produced most object questions in the Case condition (57%). In the Agr condition, the amount of object questions is lower (44%). In the AgrCa condition, the percentage of object questions is roughly as high as in the Case condition (56%).

   For the analysis, we were interested in whether speakers took into account the listener's perspective and thus cue availability. We conducted a GLMER model (see Analysis section for details) with a binomial dependent variable called question (object vs. passive question) and participant and item as random intercept factors. Group and cue were fixed factors that improved the model. An interaction between group and cue revealed that children's production of patient-initial items differed from adults' for all three cue conditions (see Table 8). For children, a significant difference was found between questions with unambiguous case (Case) and ambiguous case (Agr; $\beta = 1.84$, $z = 3.704$, $p < .01$) and between questions with unambiguous case (Case) and questions with ambiguous case on the first NP (AgrCa; $\beta = 1.58$, $z = 3.308$, $p < .01$). The distribution between object and passive
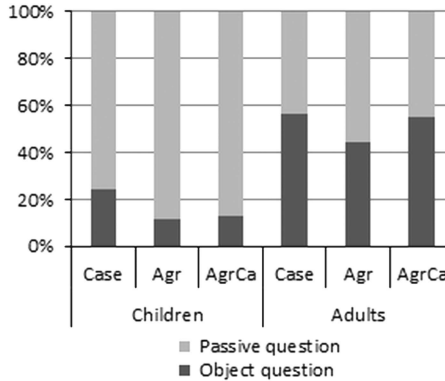
Figure 8. Distribution of object and passive *which*-questions for patient-initial items per cue condition produced by children and adults.

Table 8. *Multiple comparisons of mean accuracy of the interaction between group and the three different types of cues (Tukey contrasts)*

|  | Estimate | Standard error | z value | p value |
|---|---|---|---|---|
| Children *AgrCa*-Adults *AgrCa* | –4.8545 | 1.1124 | –4.364 | <.001*** |
| Adults *Agr*-Adults *AgrCa* | –1.0658 | 0.4173 | –2.554 | .08456† |
| Adults *Case*-Adults *AgrCa* | 0.1636 | 0.4047 | 0.404 | .99817 |
| Children *Agr*-Children *AgrCa* | –0.2646 | 0.5152 | –0.514 | .9943 |
| Children *Case*-Children *AgrCa* | 1.5766 | 0.4767 | 3.308 | .00884** |
| Children *Agr*-Adults *Agr* | –4.0533 | 1.1085 | –3.657 | .0026** |
| Adults *Case*-Adults *Agr* | 1.2294 | 0.421 | 2.92 | .03047* |
| Children *Case*-Children *Agr* | 1.8413 | 0.4971 | 3.704 | .00205** |
| Children *Case*-Adults *Case* | –3.4414 | 1.0624 | –3.239 | .01117* |

*p < .05. **p < .01. ***p < .001. †p <0.1.

questions produced by adults differed significantly between questions with unambiguous case (Case) and with ambiguous case (Agr; $\beta = 1.23$, $z = 2.92$, $p < .05$).

  To summarize, the production data reveals that children produce significantly more passive questions than adults. Moreover, both children and adults produce significantly more object questions when case can be used (either immediately or later) as a disambiguation cue by the listener than when case cannot be used as a cue.

## DISCUSSION

The aims of the study were to find out (a) whether German children and adults make use of morphosyntactic cues (case marking and verb agreement) for the comprehension of *which*-questions, (b) how these questions are processed, and (c) whether

the presence and position of morphosyntactic cues available for the listener influence the speaker's production of *which*-questions. In order to answer these questions, we first discuss children's and adults' final interpretations of *which*-questions in terms of accuracy. Then, we discuss their online processing by examining their gaze patterns. Finally, we discuss their production of *which*-questions.

### Use of morphosyntactic cues in the final interpretation of which-questions

As expected, adults correctly interpreted subject, object, and passive questions. Nevertheless, adults' accuracy scores on object questions were significantly lower than on passive questions. This suggests that object questions are more difficult than passive questions (Contemori & Belletti, 2014).

We hypothesized that some children do not show use of case and verb agreement, as they have not yet acquired the adult constraint ranking: this would affect their interpretation of object questions in that these are expected to be interpreted as subject questions. Note that similar effects are predicted by the AFH and RM account, as discussed above. Three out of 34 children consistently interpreted object questions as subject questions. Their performance supports the idea that the ranking of constraints of these children deviates from the adults' ranking by giving too much importance to the AGENT-FIRST constraint at the expense of the CASE or AGREEMENT constraints. Six more children interpreted object questions incorrectly as subject questions in half of the items. These children, who were mainly the youngest children, may still be in the process of reranking the constraints toward an adultlike ranking.

None of the children had problems interpreting passive questions. This suggests that our assumption that passive questions are easier than object questions is right, even though both are noncanonical and in both the first NP is not the agent. Why precisely this is the case remains an open question. It may be due to the more explicit morphological information, or to the structural difference between the types of noncanonicity.

Most children and all adults made use of case and verb agreement when interpreting *which*-questions. Although we did not find a significant difference in accuracy with respect to whether the object questions were disambiguated by case or by verb agreement, children's individual accuracy patterns indicate that case is a more effective disambiguation cue than verb agreement: four children made use of case, but not of verb agreement. No child showed the opposite pattern. This is in line with results from previous research on the acquisition of German relative clauses (Arosio et al., 2012) and the processing of *wh*-questions (Meng & Bader, 2000), but in contrast with findings from individuals with aphasia who showed more deficits in the processing of case cues than verb agreement cues (Hanne, Burchert, De Bleser, & Vasishth, 2015). The better performance with case cues compared to verb agreement cues may be due to the fact that case marking is directly (locally) marked on the NP, whereas agreement is marked indirectly on the verb and NP (Clahsen, 1986). For the latter, both number marking on the NP and number marking on the verb have to be recognized and linked to each other. For case, only one cue needs to be recognized, which may be easier.

The influence of working memory on object question comprehension is in line with previous studies that found that children with higher working memory capacity perform better on the comprehension of object relative clauses (e.g., Arosio et al., 2012; Booth, MacWhinney, & Harasaki, 2000; Friederici et al., 1998) and that children with low working memory showed non-adultlike attachment preferences in relative clause attachment (Felser, Marinis, & Clahsen, 2003). The finding that 8- and 9-year-old children score significantly better on object questions than 7-year-olds confirms that the comprehension of object questions develops late (and for some children not before the age of 8) in German.

### Use of morphosyntactic cues in the online processing of which-questions

The gaze data provide an answer to the question how children and adults process morphosyntactic cues in *which*-questions. As predicted by our model, adults showed no incorrect initial interpretations for subject questions and object questions disambiguated by overt case marking directly at the first NP. These types of questions were disambiguated at the beginning of the sentence, as indicated by their first and continued looks toward the target picture. In contrast, when adults processed passive questions or object questions disambiguated later in the sentence by verb agreement, they looked more toward the competitor picture, indicating that they initially interpreted these questions as subject questions. Only later did they switch interpretation, as shown by their increased looks toward the target picture.

The gaze data in this study clearly illustrates the so-called garden-path effect in adults: the adult listener is initially led to the wrong interpretation, as the literature on syntactic parsing points out (see Frazier & Clifton, 1996, for further references). As discussed above, this is explicitly predicted by the OT model, whereas other models either do not directly make this prediction (RM) or are equivocal about whether sentence revision effects are predicted or not (CM).

Based on their gaze data, children also appear to initially interpret passive questions and object questions disambiguated by verb agreement incorrectly, interpreting these questions as subject questions. Object questions disambiguated by case on the first NP did not seem to lead to incorrect initial interpretations, as children looked more toward the target picture than to the competitor picture for this type of question, although this preference increased at a later moment in time and at a slower rate than for adults. This online pattern contrasts with the children's offline responses, which did not show differences with respect to disambiguation cue. Possibly, children's gaze data, being a more sensitive measurement than offline responses, is an indication that the CASE constraint is in the process of being ranked above the AGENT-FIRST constraint (as predicted by the OT explanation).

Unlike the adults, children also showed a difference between object questions disambiguated by agreement only and object questions disambiguated by agreement and additional case marking on the second NP. The latter pattern shows a less prominent increase of looks and a quicker increase of looks toward the target picture compared to object questions solely disambiguated by agreement. This indicates that children more easily revise an incorrect initial

interpretation when an additional case cue is present. The fact that we do not find such an effect for adults does not mean that they ignore case on the second NP. Apparently, for adults the agreement cue is already sufficient evidence to revise their interpretation immediately.

Incorrect initial interpretations are found not only in the processing of object questions but also in the processing of passive questions, for both children and adults. As expected, the first NP in passive questions is interpreted as the agent, leading to first looks toward the competitor picture. For adults, switches to the target picture and thus revisions of the initial interpretation with passive questions occur earlier than with object questions. Note that even though passive questions are initially interpreted incorrectly, the accuracy scores for both children and adults are at ceiling. A revision to the correct interpretation therefore seems easier in passives than in object questions. The fact that children have to revise their interpretation for passive questions, and do so without any problems, strengthens the idea that the problems children encounter in object questions are not due to their inability to revise a first interpretation (unlike younger children in previous studies, e.g., Choi & Trueswell, 2010). Rather, the cues in passive questions might be more effective than in object questions. The *by*-agent (although not obligatory) combined with the verb form *werden* clearly indicates that the first NP is the not the agent but the patient.

The processing explanation (AFH) postulates that children's misinterpretations of object questions are due to their inability to revise an initial interpretation (Avrutin, 2000; Deevy & Leonard, 2004; Metz et al., 2010). The processing account therefore predicts similar misinterpretations in passive and object questions. Children's garden-path effect in passive questions combined with their high accuracy scores for passive questions argues against a pure processing explanation for children's misinterpretations of object questions. The RM explanation predicts correct performance regarding children's interpretation of passive questions. According to this explanation, children's misinterpretations of object questions are caused by an intervener present in object questions. Belletti (2011) argues that in passive questions there is no intervener but, instead, the movement of the object to the subject position is brought about by movement of the entire verb phrase that includes the object. This syntactic explanation therefore explains children's correct final interpretations of passive questions. However, it is unclear how it would explain the garden-path effect found in children's gaze data. In contrast, the garden-path effect in the gaze data follows straightforwardly from the incremental OT explanation.

Unlike the offline accuracy scores, the online gaze data did reveal differences with respect to cue. The question whether the cue differences are due to cue particularities (i.e., whether there is a difference between case and verb agreement, in terms of effectiveness or directness) or due to timing differences (the case cue appeared earlier in the sentence than the verb agreement cue) cannot be answered in the current study. Timing differences of the same (case) cue, which was either on the sentence-initial *wh*-phrase or on the article of the second NP, have been previously found for German children, resulting in better accuracy scores for earlier disambiguated *which*-questions (Roesch & Chondrogianni,

2016). Our finding that children's garden-path effect is less strong in object questions with verb agreement and case disambiguation on the second NP than in object questions with only verb agreement disambiguation may indicate that case is a more effective cue for children than agreement. Nevertheless, this difference can also be explained in terms of number of cues or the timing of the second cue. Questions disambiguated by two cues may be easier to process than those disambiguated by only one cue. Alternatively, the timing of the second cue may help a child who is still processing the first cue.

### Use of morphosyntactic cues in the production of which-questions

Adults' productions of patient-initial questions were clearly affected by cue. When case cues were available, significantly more object questions were produced than passive questions. When only agreement cues were available, more passive questions were produced than object questions. These are cues for the listener and not for the speaker. This indicates that adult speakers take into account the listener's ease of understanding the question, in line with the predictions of the constraint-based OT account. Note that the differences in the numbers of produced object and passive questions are significant but smaller than one may expect based on the OT account. This could be due to the fact that the participants were presented with all three cue conditions and tend to produce the same structure consistently throughout the experiment. Presenting the participants with only one condition could have led to greater differences between the numbers of produced object and passive questions.

   Children overall produced more passive questions than object questions. Nevertheless, children produced more object questions when case cues were available, compared to when case cues were unavailable. Their strong preference for passives in production seems to reflect their better understanding of passives compared to object questions and is found in previous studies as well (e.g., Armon-Lotem et al., 2016, Jensen de López et al., 2014). It has been argued that 6-year-olds still have difficulty taking into account the other person's perspective when determining the optimal form or meaning (de Hoop & Krämer, 2005/2006; Hendriks & Spenader, 2006). In order to take into account the other's perspective, children must possess sufficient theory of mind abilities. Second-order theory of mind (i.e., the ability to make inferences about someone's belief about another person's belief) develops around the age of 6 (e.g., Perner & Wimmer, 1985) and has been shown to be related to perspective taking in language (e.g., Kuijper, Hartman, & Hendriks, 2015). The results of our study suggest that the children in our study (aged 7–10) are able to take into account the listener when choosing between two question forms, producing the form that is easiest to understand for a potential listener.

   To conclude, our research shows that overall German children from 7 to 10 years old and adults make use of morphosyntactic cues such as case and verb agreement in their comprehension of *which*-questions. The gaze data show that different disambiguation cues are used at different moments in sentence processing. As predicted, our gaze data show that a revision of interpretation is

required not only when processing object questions but also when processing passive questions. These findings support the constraint-based OT account. Furthermore, children's and adults' production of *which*-questions is affected by the morphosyntactic cues that are available for the listener, indicating that 7- to 10-year-old children and adult speakers take into account potential listeners.

## APPENDIX A

TRANSLATED INSTRUCTIONS OF THE COMPREHENSION TASK

This game consists of various pictures and sentences. Each time, there are two pictures and one sentence. You have to decide which picture fits the sentence. If the picture on the left fits the sentence, then press the button on the left. If the picture on the right fits the sentence, then press the button on the right. You can press as soon as you know the answer.

TRANSLATED INSTRUCTIONS OF THE PRODUCTION TASK

In this game you see pictures in which something is missing. If you ask the right question you will see what is hidden behind the spot. The question always has to start with WELCH....

I will show you some examples to practice.

## APPENDIX B

GAZE DATA AND GAMMs: CHILDREN VERSUS ADULTS PER QUESTION TYPE

Table B.1. *Model of gaze data per different type of question (subject, object, or passive) per group (1 = children, 2 = adults)*

| A. Parametric coefficients | Estimate | Std. Error | *t* value | *p* value |
|---|---|---|---|---|
| (Intercept) | 12.270 | 1.195 | 10.272 | < .001 |
| QuestionGroupobject.1 | −9.843 | 1.683 | −5.849 | < .001 |
| QuestionGrouppassive.1 | −3.978 | 1.694 | −2.349 | < .05 |
| QuestionGroupsubject.2 | 1.358 | 1.755 | 0.774 | .4391 |
| QuestionGroupobject.2 | −3.424 | 1.743 | −1.965 | < .05 |
| QuestionGrouppassive.2 | 1.685 | 1.755 | 0.960 | .3369 |
| **B. Smooth terms** | **edf** | **Ref. *df*** | ***F* value** | ***p* value** |
| s(Timebin200):QuestionGroupsubject.1 | 5.691 | 6.733 | 18.770 | < .001 |
| s(Timebin200):QuestionGroupobject.1 | 7.022 | 8.003 | 15.658 | < .001 |
| s(Timebin200):QuestionGrouppassive.1 | 7.670 | 8.493 | 22.675 | < .001 |
| s(Timebin200):QuestionGroupsubject.2 | 6.530 | 7.573 | 15.999 | < .001 |
| s(Timebin200):QuestionGroupobject.2 | 7.224 | 8.166 | 21.723 | < .001 |
| s(Timebin200):QuestionGrouppassive.2 | 7.264 | 8.205 | 23.258 | < .001 |
| s(Timebin200,ParticipantQuestion) | 526.034 | 1722.000 | 0.976 | < .001 |

*m5c = bam(TCDiff ~ s(Timebin200,by = QuestionGroup) + QuestionGroup + s (Timebin200,ParticipantQuestion, bs = 'fs',m = 1), data = data1, gc.level = 2, method = 'ML'). Note that this summary only tells you whether the smooth of each variable is linear or not.*

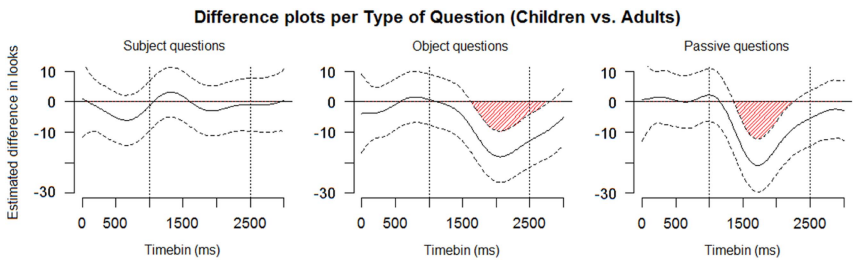**Difference plots per Type of Question (Children vs. Adults)**



Figure B.1. Difference plots per type of question (subject, object, and passive questions). The solid line represents mean; dashed lines represent upper and lower limits of the 99% confidence interval. An area indicated with red means a significant difference between children's and adults' gaze.

## APPENDIX C

GAZE DATA AND GAMMs: CHILDREN VS. ADULTS PER DISAMBIGUATION CUE

Table C.1. *Model of gaze data per different type of cue (Case, Agr, AgrCa) for object questions per group (1 = children, 2 = adults).*

| A. Parametric coefficients | Estimate | Std. error | *t* value | *p* value |
|---|---|---|---|---|
| (Intercept) | 7.182 | 1.352 | 5.309 | <.001 |
| CueGroupAgr.1 | −6.662 | 1.297 | −5.134 | <.001 |
| CueGroupAgrCa.1 | −6.181 | 1.299 | −4.757 | <.001 |
| CueGroupCase.2 | 3.919 | 1.930 | 2.030 | <.05 |
| CueGroupAgr.2 | 1.160 | 1.908 | 0.608 | .5431 |
| CueGroupAgrCa.2 | −0.3177 | 1.914 | −0.166 | .8682 |
| **B. Smooth terms** | **edf** | **Ref. *df*** | ***F* value** | ***p* value** |
| s(Timebin200) | 5.922 | 6.902 | 3.736 | <.001 |
| s(Timebin200):CueGroupCase.1 | 1.002 | 1.003 | 1.128 | .287766 |
| s(Timebin200):CueGroupAgr.1 | 6.246 | 7.350 | 12.461 | <.001 |
| s(Timebin200):CueGroupAgrCa.1 | 4.011 | 4.917 | 6.209 | <.001 |
| s(Timebin200):CueGroupCase.2 | 1.001 | 1.002 | 1.192 | .275286 |
| s(Timebin200):CueGroupAgr.2 | 6.499 | 7.585 | 10.936 | <.001 |
| s(Timebin200):CueGroupAgrCa.2 | 5.138 | 6.242 | 11.262 | <.001 |
| s(Timebin200,Participant) | 184.994 | 574.000 | 1.091 | <.001 |

*m6 = bam(TCDiff ~ s(Timebin200) + s(Timebin200,by = CueGroup) + CueGroup + s(Timebin200,Subject, bs = 'fs',m = 1), data = dataObj, gc.level = 2, method = 'ML').*

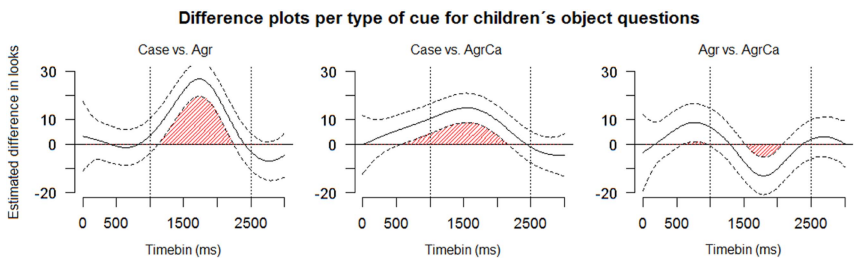**Difference plots per type of cue for children´s object questions**



Figure C.1. Difference plots per comparison type of cue (Case, Agr, and AgrCa,) for children. The solid line represents mean; dashed lines represent upper and lower limits of the 99% confidence interval. An area indicated with red means a significant difference between the two cues compared.

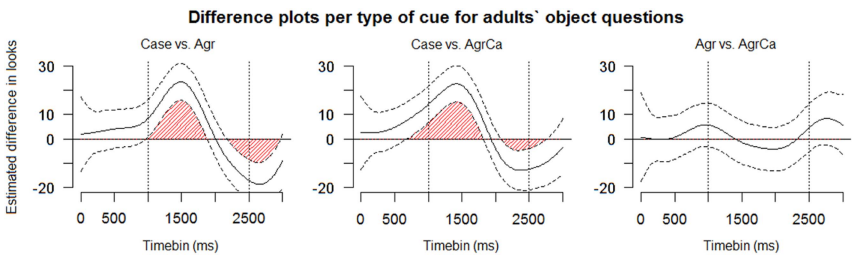**Difference plots per type of cue for adults` object questions**



Figure C.2. Difference plots per comparison type of cue (Case, Agr, and AgrCa) for adults. The solid line represents mean; dashed lines represent upper and lower limits of the 99% confidence interval. An area indicated with red means a significant difference between the two cues compared.

## NOTES

1.  This holds for the sentences in our task. The German verb *werden* can also be used as a copula or modal verb (*Welche Schülerin wird sechszehn/wird feiern?* "Which pupil will turn sixteen/will party?").
2.  A smooth function is a weighted sum of complex nonlinear basis functions (see Wood, 2006).

## REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Armon-Lotem, S., Hamns, E., Jensen de López, K., Smoczynska, M., Yatsushiro, K., Szczerbinski, M., … van der Lely, H. (2016). A large-scale cross-linguistic investigation of the acquisition of passive. *Language Acquisition*, *23*, 1.

Arosio, F., Yatsushiro, K., Forgiarini, M., & Guasti, M. (2012). Morphological information and memory resources in children's processing of relative clauses in German. *Language Learning and Development*, *8*, 340–364.

Aschermann, E., Gülzow, I., & Wendt, D. (2004). Differences in the comprehension of passive voice in German- and English-speaking children. *Swiss Journal of Psychology*, *63*, 235–245. doi:10.1024/1421-0185.63.4.235.

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.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*, 189–208. doi:10.1016/S0021-9924(03)00019-4.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. *ArXiv*, *1506*, 04967.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3–73). New York: Cambridge University Press.

Belletti, A. (2011). Considering the complexity of relative clauses and passive from the Italian perspective. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected proceedings of the Romance Turn IV*. Newcastle: Cambridge Scholars Press.

Belletti, A., & Contemori, C. (2010). Intervention and attraction. On the production of subject and object relatives by Italian (young) children and adults. In J. Costa, A. Castro, M. Lobo, & F. Pratas (Eds.), *Language acquisition and development: Generative Approaches to Language Acquisition, Proceedings of Gala 2009* (pp. 39–52). Cambridge: Cambridge Scholars Press.

Biran, M., & Ruigendijk, E. (2015). Do case and gender information assist sentence comprehension and repetition for German- and Hebrew-speaking children? *Lingua*, *164*, (Pt. A), 215–238.

Booth, J. R., MacWhinney, B., & Harasaki, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. *Child Development*, *71*, 981–1003.

Borer, H., & Wexler, K. (1987). The maturation of syntax. In T. Roeper & E. Williams (Eds.), *Parameter setting* (pp. 123–172). Dordrecht: Springer Netherlands.

Bouma, G. (2008). *Starting a sentence in Dutch: A corpus study of subject- and object-fronting* (Unpublished PhD dissertation, University of Groningen).

Brandt-Kobele, O., & Höhle, B. (2014). The detection of subject-verb agreement violations by German-speaking children: An eye-tracking study. *Lingua*, *144*, 7–20. doi:10.1016/j.lingua.2013.12.008.

Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*, 41–61.

Clahsen, H. (1986). Verb inflections in German child language: Acquisition of agreement markings and the functions they encode. *Linguistics*, *24*, 79–121. doi:10.1515/ling.1986.24.1.79.

Collins, C. (2005). A smuggling approach to the passive in English. *Syntax*, *8*, 81–120.

Contemori, C., & Belletti, A. (2014). Relatives and passive object relatives in Italian-speaking children and adults: Intervention in production and comprehension. *Applied Psycholinguistics*, *35*, 1021–1053. doi:10.1017/S0142716412000689.

Deevy, P., & Leonard, L. B. (2004). The comprehension of wh-questions in children with specific language impairment. *Journal of Speech, Language & Hearing Research*, *47*, 802–815. doi:10.1044/1092-4388(2004/060).

de Hoop, H., & Krämer, I. (2005/2006). Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition*, *13*, 103–123.

de Hoop, H., & Lamers, M. (2006). Incremental distinguishability of subject and object. In L. Kulikov, A. Malchukov & P. Swart (Eds.), *Case, valency and transitivity* (pp. 269–287). Amsterdam: Benjamins.

De Vincenzi, M., Arduino, L. S., Ciccarelli, L., & Job, R. (1999). Parsing strategies in children comprehension of interrogative sentences. In S. Bagnara (Ed.), *Proceedings of the European Conference on Cognitive Science* (pp. 301–308). Rome: Istituto di Psicologia del CNR.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, *79*, 1152–1167. doi:10.1111/j.1467-8624.2008.01181.x.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition*, *11*, 127–163.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*, 250–272.

Fikkert, P., & de Hoop, H. (2009). Language acquisition in optimality theory. *Linguistics*, *47*, 311–357. doi:10.1515/LING.2009.012.

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

Frazier, L., & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in dutch. *Journal of Memory and Language*, *28*, 331–344. doi:10.1016/0749-596X(89)90037-5.

Friederici, A. D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology*, *47*, 193–221. doi:10.1016/S0301-0511(97)00033-1.

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, *119*, 67–88. doi:10.1016/j.lingua.2008.09.002.

Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand? The comprehension of Wh questions in three subtypes of SLI. *Lingua*, *121*, 367–382. doi:10.1016/j.lingua.2010.10.004.

Guasti, M. T., Branchini, C., & Arosio, F. (2012). Interference in the production of Italian subject and object wh-questions. *Applied Psycholinguistics*, *33*, 185–223. doi:10.1017/S0142716411000324.

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.

Hendriks, P. (2014). *Asymmetries between language production and comprehension*. Dordrecht: Springer.

Hendriks, P., & Spenader, J. (2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition*, *13*, 319–348.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50* 346–363. doi:10.1002/bimj.200810425.

Jakubowicz, C. (2011). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua*, *121*, 339–351 doi:10.1016/j.lingua.2010.10.006.

Jensen de López, K., Sundahl Olsen, L., & Chondrogianni, V. (2014). Annoying Danish relatives: Comprehension and production of relative clauses in Danish typically developing children and children with SLI. *Journal of Child Language*, *41*, 51–83.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, *20*. doi:10.1017/S0272263198004045.

Kuijper, S. J. M., Hartman, C. A., & Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PLOS ONE*, *10*, e0132408.

Lindner, K. (2003). The development of sentence-interpretation strategies in monolingual German-learning children with and without specific language impairment. *Linguistics*, *41*, 213–254. doi:10.1515/ling.2003.008.

Macwhinney, B. (2005). Extending the competition model. *International Journal of Bilingualism*, *9*, 69–84. doi:10.1177/13670069050090010501.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150. doi:10.1016/S0022-5371(84)90093-8.

Maratsos, M., Fox, D. E. C., Becker, J. A., & Chalkley, M. A. (1985). Semantic restrictions on children's passives. *Cognition*, *19*, 167–191. doi:10.1016/0010-0277(85)90017-4.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. doi:10.1016/j.jml.2017.01.001.

Meng, M., & Bader, M. (2000). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech*, *43*, 43–74.

Metz, M., van Hout, A., & van der Lely, H. (2010). Understanding who and which questions in five- to nine-year-old Dutch children: The role of number. *New Developments in the Acquisition of Dutch, Groninger Arbeiten zur germanischen Linguistik*, *51*, 27–41.

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103–125. doi:10.1016/j.jml.2016.03.005.

O'Grady, W. (1997). *Syntactic development*. Chicago: University of Chicago Press.

Perner, J., & Wimmer, H. (1985). John thinks that Mary thinks that: Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *60*, 689–700.

Petermann, F., & Petermann, U. (2007). *Hamburg Wechsler Intelligenztest für Kinder—IV*. Übersetzung und Adaptation der WISC-IV® von David Wechsler. Bern: Huber.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Wiley-Blackwell.

Rizzi, L. (1990). *Relativized minimality. Linguistic Inquiry Monographs 16*. Cambridge, MA: MIT Press.

Rizzi, L. (2004). Locality and left periphery. In A. Belletti (Ed.), *Structures and beyond: The cartographic syntactic structure* (Vol. 2, pp. 223–251). Oxford: Oxford University Press.

Roesch, A. D., & Chondrogianni, V. (2015). The use of case in the comprehension of wh-questions in German-speaking children with and without SLI. In C. Hamann & E. Ruigendijk (Eds.), *Proceedings of GALA 2013* (pp. 379–402). Cambridge: Cambridge Scholars.

Roesch, A. D., & Chondrogianni, V. (2016). "Which mouse kissed the frog?": Effects of age of onset, length of exposure and knowledge of case-marking on the comprehension of wh-questions in German-speaking simultaneous and early sequential bilingual children. *Journal of Child Language*, *43*, 635–661.

Schipke, C. S., Knoll, L. J., Friederici, A. D., & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: Neural correlates of their behavioral performance. *Developmental Science*, *15*, 762–774. doi:10.1111/j.1467-7687.2012.01167.x.

Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous Wh-questions in German. In B. Hemforth & L. Konieszny (Eds.), *German sentence processing* (pp. 65–93). Dordrecht: Kluwer.

Schouwenaars, A., van Hout, A., & Hendriks, P. (2014). Word order overrules number agreement: Dutch children's interpretation and production of which-questions. In C-Y. Chu, C. E. Coughlin, B. Lopez Prego, U. Minai, & A. Tremblay (Eds.), *Selected Proceedings of the 5th Conference on Generative Approaches to Language Acquisition North America (GALANA 2012)* (pp. 60–71). Somerville, MA: Cascadilla Proceedings Project.

Seymour, H., Roeper, T., & de Villiers, J. (2003). *DELV-ST (Diagnostic Evaluation of Language Variation) Screening Test*. San Antonio, TX: Psychological Corporation.

Smolensky, P. (1996). *The initial state and "richness of the base" in optimality theory* (Tech. Rep. No. JHU-CogSci-96-4). Baltimore, MD: Johns Hopkins University, Cognitive Science Department.

Stevenson, S., & Smolensky, P. (2006). Optimality in sentence processing. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind*. Cambridge, MA: MIT Press.

Stromswold, K. (1995). The acquisition of subject and object Wh-questions. *Language Acquisition*, *4*, 5–48. doi:10.1080/10489223.1995.9671658.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, *30*, 395–402.

van Rij, J., Baayen, R. H., Wieling, M., & van Rijn, H. (2015). itsadug: Interpreting time series, autocorrelated data using GAMMs. R package version 1.0.4. Vienna, Austria: R Core Team.

van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (Linguistische Arbeiten, Vol. 563, pp. 267–293). Berlin: De Gruyter.

Wood, S. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: CRC Press.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Part B*, *73*, 3–36.

Zeevat, H. (2006). Freezing and marking. *Linguistics*, *44*, 1095–1111. doi:10.1515/LING.2006.035.