Title

The real-time comprehension of WH-dependencies in a WH-agreement language

Journal

Language, 91(1)

ISSN

0097-8507

Authors

Wagers, Matthew

Borja, Manuel F

Chung, Sandra

Publication Date

2015-03-01

DOI

10.1353/lan.2015.0001

Peer reviewed

The real-time comprehension of wh-dependencies in a Wh-Agreement language

Matthew Wagers†
University of California, Santa Cruz
mwagers@ucsc.edu

Manuel F. Borja
Inetnun Åmut yan Kutturan Natibu
ianasfnma@gmail.com

Sandra Chung
University of California, Santa Cruz
schung@ucsc.edu

† University of California, Santa Cruz
Department of Linguistics
1156 High St
Santa Cruz, CA 95064

The real-time comprehension of wh-dependencies in a Wh-Agreement language

ABSTRACT

In the verb-initial language Chamorro, an Austronesian language of the Mariana Islands, wh-dependencies exhibit a special verbal inflection known as Wh-Agreement: verbs along the path of the wh-dependency are inflected for the grammatical relation of the gap and the intermediate landing sites of the filler. Two on-line comprehension experiments conducted in the Northern Mariana Islands reveal that the morphological paradigm of Wh-Agreement affects the timing of dependency formation and interpretation in this language. Overt Wh-Agreement facilitates the formation of a wh-dependency. However, when overt Wh-Agreement could occur but does not, its absence delays and attenuates wh-dependency formation. In short, morphological information exerts a powerful influence on the unfolding parse, one which has temporal priority over syntactic information, such as word order, and semantic information, such as argument structure.*

Keywords: wh-dependencies, comprehension, Chamorro, Wh-Agreement, possessors, Active Filler Strategy

1. INTRODUCTION. Much research in psycholinguistics has been devoted to the study of the real-time comprehension of wh-dependencies—the dependencies characteristic of relative clauses, constituent questions, comparatives, focus constructions, and the like. In a wh-dependency, a constituent (the FILLER) appears displaced from the site in which it would normally have been expected to occur (the GAP). The filler is in italics and the gap is marked with an underline in the examples below.[1]

(1)a. The student *who* Marisa might talk to __ tomorrow will be nervous.

b. *Which student* do you think Marisa might talk to __ tomorrow?

c. ?Joe knows the girl *who* for the two of us to talk to her tomorrow would be a mistake.

d. ?*Which girl* do you think that the two of us might talk to her and one other person tomorrow?

Several factors conspire to make the comprehension of wh-dependencies a prime target for psycholinguistic investigation. To begin with, the distance between the filler and the gap is in principle unbounded, both linearly and in terms of hierarchical structure (Ross 1967, McElree et al. 2003). Second, the dependency is nonetheless sensitive to structural constraints, such as island constraints (Ross 1967, Phillips 2006). Third, the gap itself is not flagged unequivocally in the linguistic signal. Instead, it is indicated ambiguously (Fodor 1978), by the absence of a constituent (see 1a-1b) or by an ordinary-looking pronoun that stands in for the filler (a resumptive pronoun; 1c-1d). All this could potentially leave the comprehender in a state of uncertainty throughout the course of the wh-dependency about the filler's grammatical relation and semantic role .

What types of linguistic knowledge have the potential to ease that uncertainty and contribute to the dependency's resolution? Here we investigate this question with respect to the morphology of extraction found in Chamorro, an Austronesian language spoken in the Mariana Islands (in Micronesia). In Chamorro, wh-dependencies exhibit a special verb agreement known as Wh-Agreement: the verbs along the path of the wh-dependency are inflected for the grammatical relation of the gap and the filler's intermediate landing sites (Chung 1982, 1998). The specific goal of this study is to determine whether Wh-Agreement in Chamorro has a functional impact on the real-time comprehension of wh-dependencies. We ask whether, and to what extent, comprehenders use the cues provided by the

agreement morphology to constrain their real-time interpretation of a Chamorro sentence as it unfolds. In particular, we ask whether this morphological information interacts with a general strategy, observed in other languages, that impels comprehenders to predictively link fillers to gaps, i.e., the Active Filler Strategy (Frazier 1987).

Over and above this, our investigation has two larger, interconnected aims. First, to contribute to diversifying the language data that inform the construction of psycholinguistic theory. In general, research on sentence processing has drawn from an extremely small, skewed sample of Western European and East Asian languages (Hawkins 2007, Jaeger & Norcliffe 2009, Anand et al. 2011). Anand and colleagues (2011) estimated that just ten languages account for at least 85% of the research reported in psycholinguistics journals and conferences: these are, in rank order, English, German, Japanese, French, Dutch, Spanish, Chinese, Korean, Finnish and Italian. These languages represent a very narrow slice of the world's linguistic diversity. Virtually absent from psycholinguistic research is data from the Austronesian language family, which is the world's second largest in terms of number of distinct languages (Lewis et al. 2013). In view of this, it is hard to know to what extent current psycholinguistic theory is a general theory of human capacity and to what extent its claims might be skewed by the handful of languages studied to date. Our research is part of a nascent but growing effort to incorporate data from understudied languages into psycholinguistic research (what Christianson (2002) dubs 'field psycholinguistics'; see e.g. Christianson & Ferreira 2005 and Christianson & Cho 2009 on Odawa; Harris & Samuel 2011 on Batsbi; Clemens et al. 2014 on Ch'ol and Q'anjob'al; and Gagliardi & Lidz 2013 on Tsez). Much of this research uses controlled experimental designs to characterize off-line patterns of acceptability, use, and interpretation that are the end result of comprehension or production. The present project seeks to extend this research by characterizing patterns of complexity and interpretation during real-time comprehension, i.e., on-line, in the midst of processing an expression.

Linguistic diversity is not the only issue at stake, however. The vast majority of the world's languages are small languages, spoken by populations of fewer than a million speakers. These populations typically lack socio-economic influence, are not literate, and do not share Western cultural presuppositions. The speakers of most of the world's

languages are also not the typical participants in the experimental studies that shape cognitive science (Henrich et al. 2010). The methodologies of experimental studies are closely adapted to certain first-world cultural conventions. They presuppose compliance with the experimenter's authority, willingness to engage in arbitrary tasks for long periods of time, and the appeal of incentives that target the individual as opposed to the group (Rosenthal & Rosnow 2009). Additionally, the design and analysis of such studies often require large amounts of data to be collected. Many of these cultural conventions are absent when the participants in experimental research are speakers of a smaller, non-Western language, such as Chamorro. Thus, our second larger aim was to understand how we could minimize the first-world design features of our experiments but still achieve reliable scientific results. Our research team, which consists of a laboratory psycholinguist, a native-speaker educator, and a syntactician-fieldworker, was formed expressly to tackle this question.

In the rest of this section, we give a brief sketch of Wh-Agreement in Chamorro, situate our investigation in the context of previous research on the comprehension of wh-dependencies, and describe the specific questions and design features of our study.

1.1. WH-AGREEMENT IN CHAMORRO. Chamorro is a head-initial language in which the predicate comes first in the clause. When the predicate is a verb or adjective, the neutral word order is *Verb Subject Object Other*, where *Other* includes other arguments and adjuncts. (To simplify the exposition, predicate adjectives are henceforth treated as a subtype of verb.)

(2) **VSOX Word Order**

Ha	pupulan	si	Antonio i	famagu'un	gi	egga'an.
AGR	watch.PROG	UNM	Antonio the	children	LOC	morning

'Antonio is watching the children in the morning.'

However, the word order of arguments and adjuncts following the verb is not rigid, but flexible. The subject can follow the object; adjuncts can precede arguments; etc. Some of the options are shown in 3.

(3)a. **VOSX Word Order**

Ha	pupulan	i	famagu'un	si	Antonio	gi	egga'an.
AGR	watch.PROG	the	children	UNM	Antonio	LOC	morning

'Antonio is watching the children in the morning.'

b. **VSXO Word Order**

Para	u	bisita	si	Maria	agupa'	i	mås	påtgun	na
FUT	AGR	visit	UNM	Maria	tomorrow	the	most	child	L

hagå-ña.
daughter-AGR

'Maria will visit her youngest daughter tomorrow.'

The verb agrees with its subject in person and/or number; this subject-verb agreement is realized by morphemes that also encode mood and transitivity. In the realis mood, the agreement is realized as a proclitic when the verb is transitive (e.g. *ha* ' 3sg.' in 3a) but as a prefix or infix (e.g. *-um-* or no morpheme 'sg.', *man-* 'pl.'; see 13a) when the verb is intransitive.

Chamorro has null arguments. Subject pronouns can be null—in fact, they must be null if the verb agrees with them in person, as in 4. Object pronouns can be null; possessor pronouns are always null.

(4) **Pronouns that trigger agreement in person are always null**

Hu	taitai	(*yu')	i	lepblom-mu	(*hågu).
SUBJ.AGR	read	I	the	book-POSS.AGR	you

'I read your book.'

Two aspects of Chamorro's syntactic profile figure prominently in this investigation. First, constituent questions involve a wh-dependency in which the gap can bear various grammatical relations: these include subject, object, oblique argument, adjunct, possessor of the object, and possessor of an intransitive subject. Second, Chamorro has Wh-Agreement (Chung 1982, 1994, 1998). This special agreement, which replaces the normal subject-verb agreement, registers both the presence of a wh-dependency and the

grammatical relation of the gap. As is common in morphological paradigms, some forms of Wh-Agreement are overt but others are not (compare English *she think-s* with *I think*, *you think*). When the gap is the subject of a transitive verb in the realis mood, the verb shows nominative Wh-Agreement, via the infix -*um*- (see 5a). When the gap is the oblique complement of an intransitive verb, the verb shows oblique Wh-Agreement (obligatorily), via nominalization of the verb (5b).[2] When the gap is the direct object, the verb optionally shows objective Wh-Agreement, via nominalization of the verb plus the infix -*in*-. However, when the gap is a possessor, the verb does not show Wh-Agreement, but instead is inflected for normal subject-verb agreement (5c).

(5)a. **Subject gap: nominative Wh-Agreement**

Håyi fumåhan i kareta?

who? WH[SBJ].buy the car

'Who bought the car?'

b. **Oblique gap: oblique Wh-Agreement**

Håyi maleffåm-mu?

who? WH[OBL].forget-AGR

'Who did you forget?'

c. **Possessor gap: no Wh-Agreement**

Håyi un fåhan karetå-ña?

who? AGR buy car-AGR

'Whose car did you buy (lit. Who did you buy a car of)?'

In short, verbs that show Wh-Agreement carry extra information about the grammatical relation of the gap. This fact suggests that research on Chamorro could provide a novel window into the real-time comprehension of wh-dependencies.

Generally, for a verb to be inflected for Wh-Agreement, the gap must be an argument of the verb. Other features of the verb, such as transitivity and mood, are also relevant. Even when these conditions are met, Wh-Agreement is not necessarily realized overtly. Table 1 summarizes the Wh-Agreement paradigm for verbs in the realis mood. Crucially, Wh-Agreement is not overt when the gap is the subject of an intransitive verb or a possessor,

and it need not be overt when the gap is a direct object. These cases are indicated in bold face in Table 1.

<INSERT TABLE 1 HERE>

We now consider the grammar of Chamorro from the perspective of the language comprehension system and the way in which the incremental interpretation of expressions evolves over time. Specifically, we ask whether the presence of Wh-Agreement morphology, or its absence, changes the identity of the cues to the grammatical relation of the gap and their distribution across time. Does Chamorro verb morphology cause the comprehender to arrive at the correct interpretation of a wh-dependency in different ways or at different times? If so, how does sensitivity to this information interact with the other strategies that the comprehender employs? To answer these questions, let us first consider findings from other languages on the processing of wh-dependencies.

1.2. COMPREHENDING WH-DEPENDENCIES. As a consequence of Wh-Agreement, the Chamorro verb will provide an unambiguous cue to the grammatical relation of the gap in some wh-dependencies but not others. Concretely, we can compare wh-dependencies in which the gap is a direct object (henceforth, OBJECT EXTRACTION) with wh-dependencies in which the gap is a possessor of the direct object (POSSESSOR EXTRACTION), as in 6.

(6)a. **Object extraction, (transitive verb, no Wh-Agreement)**

Håfa na magågu un låksi gi paingi?
what? L clothes AGR sew LOC last.night
'Which clothes did you sew last night?'

b. **Possessor extraction (transitive verb, no Wh-Agreement)**

Håfa na magågu un låksi gi paingi manggas-ñiha?
what? L clothes AGR sew LOC last.night sleeves-AGR
'Which clothes did you sew sleeves of last night?'

c. **Object extraction (transitive verb, overt Wh-Agreement)**

Håfa na magågu l**in**aksem-**mu** gi paingi?
what? L clothes WH[OBJ].sew-AGR LOC last.night
'Which clothes did you sew last night?'

A comprehender who has heard the first five morphemes of 6a or 6b (underlined) could entertain analyses in which the gap is a direct object or a possessor. Only when the clause boundary is encountered in 6a, or the possessed DP is encountered in 6b, is the sentence fully disambiguated. In contrast, a comprehender who hears 6c, in which the verb *linaksem-mu* shows overt Wh-Agreement, can know by the verb alone that the gap must receive an object analysis.

The situation that the Chamorro comprehender confronts in processing the underlined portions of 6a-b is analogous to the situation faced by comprehenders in languages, such as English, Italian, or Dutch, in which the verb form does not covary with the presence of a wh-dependency. Thus, like a Chamorro speaker who hears 6a-b, an English comprehender can entertain alternative analyses when faced with the first words from the sentences below.

(7) a. Which book did you read last night?

b. Which book did you read a review of last night?

(8) a. Which recording of the *Matthäus Passion* do you prefer?

b. Which recording of the *Matthäus Passion* do you prefer to listen to?

Considerable effort has been devoted to understanding how comprehenders negotiate situations in which alternative analyses arrive: Do they maintain these analyses in parallel? If not, how do they choose which analysis, if any, to pursue?

Previous investigations have revealed a strong pressure to link the filler to a gap and thereby assign it a grammatical relation, even if the initial linkage turns out to be wrong and to require revision (Crain & Fodor 1985, Stowe 1986, Frazier 1987, Frazier & Flores D'Arcais 1989, Pickering & Traxler 1996, Aoshima et al. 2004). Frazier (1987) dubbed this the ACTIVE FILLER STRATEGY. The Active Filler Strategy, which has broad crosslinguistic support (Phillips & Wagers 2007), is conceivably universal. There are various potential sources for it: the burden on working memory imposed by syntactically or semantically unintegrated constituents (Wanner & Maratsos 1978, Gibson 1998, 2000; cf. Wagers 2012), a specific drive to link arguments to grammatical relations as soon as possible (Pritchett 1992; Aoshima et al. 2004), or a more general drive to derive an interpretation as soon as possible (e.g. Sedivy 1999). Crocker & Chater (1998) argue that early commitment to a particular analysis may be an optimizing strategy, because it provides a potentially

falsifiable hypothesis to test the incoming input against. On this view, it is preferable to pursue a hypothesis with a sharp, falsifiable prediction than to maintain a prolonged ambiguity. In a similar vein, a comprehender may gain an incidental benefit from an incorrect analysis if, by assigning the filler a grammatical relation, she increases its strength in memory (Wagers 2008). While the comprehender may simply make the right initial decision, this is not guaranteed. In this sense, the Active Filler Strategy represents a systematic risk that the comprehender takes in the process of interpreting partial input.

What are the ways in which a comprehender could mitigate this risk? That is, how can the comprehender minimize mistakes without delaying the analysis? One way is to consult her knowledge of the possible, impossible, likely, and preferable ways of continuing the partial expression. The sources of this knowledge are rich, coming from the combinatorial possibilities provided by the grammar, the likelihood of combinations learned by experience, the specific discourse context, etc. Because this knowledge is distributed across the linguistic system—in the lexicon, morphology, phrase structure, conceptual structure, general knowledge of the world—its use in resolving a wh-dependency is an unfolding, dynamic process: different types of information may be accessed at different times or apply with different strengths. For example, constraints on phrase structure and movement have a strong impact on the comprehension of wh-dependencies. Comprehenders show an early sensitivity to island constraints—constraints that ban a wh-dependency from linking a filler to a gap inside particular domains (Stowe 1986, Traxler & Pickering 1996, Phillips 2006). Comprehenders do not entertain analyses in which the wh-dependency terminates inside an island. They will, however, entertain analyses in which an island-crossing dependency could be licensed by an additional gap, as occurs in parasitic gap constructions (Phillips 2006) or across-the-board movement (Wagers & Phillips 2009).

Furthermore, the comprehender's incremental analysis of wh-dependencies is constrained by information contained in the verb's lexical entry, including subcategorization and argument structure (Boland et al. 1995; Pickering & Traxler 2001, 2003). Boland and colleagues (1995) investigated whether the active comprehension of wh-dependencies is affected by the verb's subcategorization possibilities; specifically, the ability to subcategorize for a second, optional complement. The verb *remind*, for instance, subcategorizes for a direct object and (optionally) for an embedded clause. In the

incomplete sentence 9a, the filler *which movie* can be analyzed in two ways: either as the direct object of *remind*, in which case the sentence would be anomalous, or as part of the embedded clause, as in 9b.

(9) a. *Which movie* did you remind ...

b. *Which movie* did you remind your boyfriend to record ___ ?

To test whether the comprehender is sensitive to these options, they used the stops-making-sense task, in which participants are instructed to end the presentation of a sentence if it "stops making sense". Comprehenders often stopped the sentence after encountering a verb that subcategorized for just one complement, but did so far less often after encountering a verb that permits two, like *remind*.

Pickering and Traxler (2001) replicated this finding with eye-tracking: implausible combinations like 9a led to elevated reading times at the verb when the verb subcategorized for just one complement, but not when it could subcategorize for two. On the other hand, Pickering and Traxler (2003) showed that the *frequency* of a verb's subcategorization possibilities does not affect the Active Filler Strategy. They investigated verbs with multiple, *exclusive* subcategorizations, such as the verb *worry*, and asked how comprehenders choose among them. For instance, the incomplete sentence 10a can continue either as 10b or 10c.

(10)a. *Which worker* did the manager worry ...

b. *Which worker* did the manager worry ___?

c. *Which worker* did the manager worry about ___ ?

Although continuations like 10c are more frequent, they found that comprehenders disregard this subcategorization and initially adopt the object analysis in 10b, as predicted by the Active Filler Strategy. In short, comprehenders rely on access to lexically-specific information supplied by the verb, and the structure of lexical entries contributes in non-trivial ways to the application of the Active Filler Strategy. The existence of multiple potential gap sites within the same subcategorization can affect where the gap is posited. But interestingly, the frequency of mutually exclusive subcategorizations does not.

Another set of cues that may contribute to the application of the Active Filler Strategy is provided by inflectional morphology, either on the filler or along the path to the gap. For example, case-marking on the filler has been shown to guide the real-time interpretation of

wh-dependencies in German (e.g. Fiebach et al. 2001). Closed-class morphology is accessed very early in word identification, probably before the perceptual stimulus is matched against the open-class lexicon (Taft & Forster 1975, Dikker et al. 2009, Solomyak & Marantz 2010, Lehtonen et al. 2011). This suggests that morphological cues could have a temporal advantage over lexically-specific cues in comprehension. In Chamorro, as we have seen, overt Wh-Agreement provides direct morphological evidence for the grammatical relation of the gap. This type of evidence is language-general in Chamorro, not lexically-specific information anchored to particular verbs. Moreover, because the information provided by overt Wh-Agreement points to a particular gap site, it contrasts with information provided by island structures, which generally only exclude certain analyses, and the lexical properties of verbs, which probabilistically promote certain analyses but rarely point uniquely to one. These considerations led us to ask whether and to what extent Wh-Agreement has an impact on the comprehension of Chamorro wh-dependencies. The question comes in two parts: First, what happens when Wh-Agreement is present and visibly encodes the link between the filler and the gap? Second, what happens when Wh-Agreement is absent? Would the comprehender then have recourse to a heuristic like the Active Filler Strategy, or would she interpret the absence of Wh-Agreement as meaningful?

1.3. THE PRESENT STUDY. Recall that in Chamorro, overt Wh-Agreement provides an unambiguous cue to the grammatical relation of the gap. The inflected verb *l<u>in</u>aksem-<u>mu</u>* (<u>'you sewed (WH[OBJ])'</u>), by virtue of the underlined Wh-Agreement morphology, provides strong and indefeasible evidence that the gap is a direct object. And, by conjecture, so does any verb with the same Wh-Agreement morphology. On the other hand, the absence of overt Wh-Agreement provides an ambiguous cue to the grammatical relation of the gap. The inflected verb *un låksi* ('you sewed') is compatible with wh-dependencies in which the gap is a direct object *or* the possessor of a direct object (as well as sentences in which there is no wh-dependency at all). A comprehender facing a verb like *un låksi* may make an English-like decision to identify the gap site with the object position; on the other hand, she may make no decisions in the absence of Wh-Agreement, either because it is overt morphology that directly guides parsing in Chamorro or because she regards it as informative that Wh-Agreement morphology was not expressed. This led us to ask whether

comprehenders make different decisions about the gap site when confronted with overt Wh-Agreement versus no Wh-Agreement.

To answer this question, we conducted two experiments with Chamorro speakers in the U.S. Commonwealth of the Northern Mariana Islands (CNMI). Both experiments were inspired by the designs in Boland et al. 1995 and Pickering & Traxler 2001, 2003. We manipulated the goodness-of-fit of a filler as an argument of the verb, and crossed this manipulation with the verb's morphological form. We obtained behavioral correlates of goodness-of-fit by using two distinct methodologies: self-paced listening and a variant of preferential looking. These methods respond to a final feature of our study: we invited the participation of a very broad demographic swath of the Chamorro community.

2. METHODS AND MATERIALS. The larger cultural context of the Chamorro language differs significantly from the contexts in which psycholinguistic experiments are typically conducted. First, although Chamorro was once the dominant language of the Mariana Islands, it is now on the cusp of language endangerment. The CNMI has two indigenous populations, Chamorros and Carolinians, as well as numerous foreign residents from elsewhere in Micronesia, the Philippines, and Asia. In this multilingual, multicultural environment, English is the language of public settings. Chamorro was the language of most Chamorro homes in the CNMI until the early 1980s, when it began to be replaced by English. Second, most speakers of Chamorro are literate in English but not in Chamorro. Literacy is made more difficult by the existence of several competing orthographies. Older Chamorros—those most likely to be fluent speakers of the language—tend to have limited formal education, limited experience with test-taking, and limited computer skills. These aspects of the cultural context posed challenges to our experimental design.

Our efforts to design an experimental study that would be culturally appropriate and would deliver accurate real-time measurements led us to make numerous departures from standard data-gathering methods. Our experiment involved listening rather than reading, anomaly judgments rather than grammaticality judgments, and was relatively short. The entire listening task took about 6 minutes to complete and was presented in two modes. Participants whose occupations involve frequent computer use—like teachers or accountants—completed the task in a self-paced listening paradigm (section 2.4). Other

participants completed the task in a preferential looking paradigm (section 2.5): they simply listened to a recording of the question while looking at a laptop screen that showed two boxes, a green box with the word *måolik* 'good' and a red box with the words *ti måolik* 'not good'. Their eye movements were recorded, with their permission, by the laptop webcam and later coded blind by multiple annotators. This simple method gave us a record of how their comprehension proceeded. At the end of each sentence, all participants, regardless of the presentation mode completed the same forced-choice task (section 2.3).

The experiment was conducted entirely in Chamorro because we felt that this would cause participants to be more highly engaged. Finally, because Chamorro culture emphasizes the involvement of all members of the community, we invited broad participation from individuals of different backgrounds and ages.

2.1. PARTICIPANTS. We recruited participants primarily via word-of-mouth on the island of Saipan and with the help of two local residents on the islands of Tinian and Rota. Participants took the experiment during July 2011; we did not begin analysis until all participants' data had been collected. Altogether data were collected from 112 participants, ranging in age from 19 to 81. In an oral interview with each participant before the experiment, we asked for age, island of birth, and how much time, if any, had been spent off-island (*sanlagu*). Figure 1 summarizes the demographic characteristics of our sample.

Consent to participate in the experiment, and to have video taken with the laptop camera, was obtained orally. An exemption from collecting written consent was granted by the UC Santa Cruz Institutional Review Board, on the grounds that requiring participants to sign a form had the potential to induce anxiety and/or discourage participation. This risk seemed non-trivial, given that no individuals we recruited had participated in a behavioral study before. For their participation in the experiment, individuals received a 4GB USB flash drive (approximate value, when purchased in the mainland US: $7).

<INSERT FIGURE 1 HERE>

2.2. MATERIALS. Our materials crossed the grammatical relation of the gap with the filler's plausibility as an argument of the verb. For transitive verbs, the gap was either the

direct object or the possessor of the direct object.[3] In OBJECT EXTRACTIONS, the verb showed overt Wh-Agreement; in POSSESSOR EXTRACTIONS, the verb occurred in its citation form, with normal subject-verb agreement (see 11-12 and Table 1). We manipulated the plausibility of the filler as the verb's internal argument. In object extractions, the filler is actually the internal argument of the verb, whereas in possessor extractions, it is not; instead, the verb's internal argument is the entire possessed DP. Thus, in object extractions, a filler that is an IMPLAUSIBLE ARGUMENT is implausible as the verb's internal argument throughout the time-course of dependency formation. But in possessor extractions, a filler that we call an IMPLAUSIBLE ARGUMENT is temporarily implausible as the verb's internal argument but ultimately plausible as a possessor. The sample item set in 11-12 illustrates the design. The critical region (underlined) begins with the onset of the verb and ends with the offset of the shared adjunct: in this region, the filler in 11 must be interpreted as the direct object, but in 12 the filler could be interpreted either as the direct object or the direct object's possessor.

(11) **Object extraction (transitive verb, overt Wh-Agreement)**

a. Plausible argument

Kuåntu na chinina <u>prinensåm-mu nigap</u> gi talu'åni?
how.many? L shirts WH[OBJ].iron-AGR yesterday LOC afternoon
'How many shirts did you iron _ yesterday afternoon?'

b. Implausible argument

Kuåntu na patgun låhi <u>prinensåm-mu nigap</u> gi talu'åni?
how.many? L child male WH[OBJ].iron-AGR yesterday LOC afternoon
'#How many boys did you iron _ yesterday afternoon?'

(12) **Possessor extraction (transitive verb, no Wh-Agreement)**

a. Plausible argument

Kuåntu na chinina <u>un prensa nigap</u> manggas-ñiha?
how.many? L shirts AGR iron yesterday sleeves-AGR
'How many shirts did you iron [sleeves of _] yesterday?'

b. Implausible argument

Kuåntu na patgun låhi <u>un prensa nigap</u> chininan-ñiha?
how.many? L child male AGR iron yesterday shirts-AGR
'How many boys did you iron [shirts of _] yesterday?'

We created 12 item sets containing transitive verbs. As a set of control conditions in which the verb's morphological form is unrevealing, we created 12 further item sets containing intransitive verbs. When the gap is the subject of an intransitive verb, the verb does not have overt Wh-Agreement (see 13-14 and Table 1). For these verbs, the gap was either the subject or the possessor of the subject. Once again, we manipulated the plausibility of the filler as an argument of the verb. In the subject extraction in 13b, the filler—an implausible argument—is implausible as the verb's subject argument throughout the time-course of dependency formation. But in the possessor extraction in 14b, the filler—an implausible argument—is temporarily implausible as the verb's subject argument but ultimately plausible as a possessor.

(13) **Subject extraction (intransitive verb, no Wh-Agreement)**

a. Plausible argument

Håfa na guma' kimason gi ma'pus na såkkan?
which? L house AGR.burned LOC last L year
'Which house _ burned down last year?'

b. Implausible argument

Håyi na doktu kimason gi ma'pus na såkkan?
who? L doctor AGR.burned LOC last L year
'#Which doctor _ burned down last year?'

(14) **Possessor extraction (intransitive verb, no Wh-Agreement)**

a. Plausible argument

Håfa na guma' kimason gi ma'pus na såkkan atof-ña?
which? L house AGR.burned LOC last L year roof-AGR
'Which house did [a roof of _] burn down last year?'

b. Implausible argument

Håyi na doktu kimason gi ma'pus na såkkan gimå'-ña?
who? L doctor AGR.burned LOC last L year house-AGR
'Which doctor did [a house of _] burn down last year?'

In addition to the 24 item sets just mentioned, we included 16 filler sentences. Four filler sentences contained a transitive subject gap (and overt Wh-Agreement); 4 contained an oblique gap (and overt Wh-Agreement); and 4 contained an adjunct gap (and no Wh-

Agreement). Six of these filler sentences were plausible and 6 were implausible. The final 4 filler sentences contained a genuinely anomalous possessor extraction that was temporarily plausible if the filler was interpreted as an argument of the verb. A sample of filler sentences is given in 15.

(15)a. **Implausible transitive subject extraction (overt Wh-Agreement)**

Håyi na guellu mu-na'susu i patgun?
who? L grandfather WH[SBJ]-suckle D child
'#Which grandfather breast-fed the child?'

b. **Plausible oblique extraction (overt Wh-Agreement)**

Håfa na påyu maleffan-ñiha gi iskuela?
what? L umbrella forget-AGR LOC school
'Which umbrella did they forget at school?'

c. **Implausible possessor extraction (no Wh-Agreement)**

Håfa na klasin kåhun ais manenghing todu i tiempu minidong-ña?
what? L kind box ice AGR.cold all D time size-AGR
'#What kind of refrigerator is [the size of_] always cold?'

d. **Plausible adjunct extraction (no Wh-Agreement)**

Månu na guma' Yu'us ni ma-po'lu i matai?
where? L house god C AGR.PASS-place D dead
'Which church was the dead person placed in?'

Across both the item sets and the filler sentences, animacy of the filler was counter-balanced with plausibility, so that animacy was never a cue to the sentence's degree of anomaly. The overall distribution of item sets and filler sentences according to extraction type (= the grammatical relation of the gap) and the filler's plausibility as an argument is given in Table 2.

<INSERT TABLE 2 HERE>

Item sets and filler sentences were read by Borja and digitally recorded with a Zoom H4N portable recorder (Zoom Corporation, Tokyo, Japan). These recordings were spliced into individual sentence files and amplitude-normalized. Each sentence file was further

segmented into four major constituents using Praat (Boersma & Weenink 2014) for the self-paced listening task (section 2.5)
Item sets and instructions, recordings, and presentation scripts are available as Supplementary Materials from the authors' website and deposited in their institutional repository.

2.3. METHODS: FORCED CHOICE TASK. In both modes of presentation, participants were presented with a response screen, shown in Figure 2, at the offset of the sentence stimulus. To the left of a centered question mark was a red box containing the words *ti måolik* ('not good'); to the right, a green box containing the word *måolik* ('good'). These colored boxes mapped to response buttons on the keyboard ('f', 'j') which were marked with red and green stickers, respectively. Participants were instructed to select *ti måolik* ('not good') if the question they just heard did not make sense, and *måolik* ('good') otherwise.

Each experimental session consisted of an instruction phase, during which participants were familiarized with the experimental equipment and instructed how to respond in the forced choice sensicality task. After six practice trials, participants completed the main study, which lasted from 6 to 10 minutes depending on the participant's speed. Finally, participants were debriefed on their impressions of the experimental stimuli and given information about the purpose of the experiment. Debriefing sessions lasted up to 30 minutes, depending on the participant's interest. The orientation and instructions were delivered in Chamorro, either orally by Borja or via an audio recording. Most debriefings occurred primarily in Chamorro, but some participants also switched to English.

<INSERT FIGURE 2 HERE>

2.4. METHODS: SELF-PACED LISTENING. In the self-paced listening task, participants listened to the stimuli in a moving window fashion (Ferreira et al. 1996) using a modified version of the Linger software package (Rohde 2003). Sentences were segmented into four major constituents: (1) the filler, (2) the verb and its agreement morphology, (3) the adjunct phrase immediately following the verb, and (4) the sentence-final constituent, which was a second adjunct in object extractions, and the possessed DP in possessor extractions. Thus,

example 11a would be segmented as follows: (1) *Kuåntu na chinina* | (2) *prinensåm-mu* | (3) *nigap* | (4) *gi talu'åni* ('(1) How many shirts | (2) did you iron | (3) yesterday | (4) in the afternoon'). These segments corresponded to phonological phrases. Participants listened to each segment over acoustically-shielded headphones and advanced to the next segment by pressing the spacebar. After participants heard the final segment, they pressed the spacebar once more to reveal the response screen.

Self-paced listening data were analyzed for segment-by-segment reaction times (RT) and for accuracy. We first trimmed the data by removing listening times below the .005 RT quantile and above the .995 RT quantile. This resulted in a maximum listening time of 7370 ms. For each segment, a linear mixed effects model was estimated with the *lme4* package (Bates et al. 2012) in the R language (R Core Team, 2013). The dependent variable was listening time per segment, and it was first transformed by the natural logarithm. We entered the experimental factors, Extraction Type (= the grammatical relation of the gap) and Argument Plausibility (of the filler), and their interaction as fixed effects. Random intercepts were estimated both for participants and for items.[4] After a model was initially estimated, we identified and removed observations whose normalized residual was greater or less than 2.5 (Baayen & Milin 2010). Both trimmed and untrimmed models are reported. In section 3.1, we discuss a further analysis in which the acoustic duration of the segment was included in the models.

Before analyzing the RTs, we applied a d-prime criterion for excluding participants who might not have performed the task according to instructions. D-prime is a measure of discriminability that corrects for response bias (MacMillan & Creelman 1991). It is calculated by taking the difference between the proportion of correct answers in one condition whose expected answer is *måolik* ('good')/'yes' (p_{Hit}) and the proportion of incorrect answers in a closely-matched corresponding condition whose expected answer is *ti måolik* ('not good')/'no' ($p_{\text{FalseAlarm}}$). Proportions are first converted into probits by applying the normal quantile function. We calculated p_{Hit} from the plausible filler sentences and $p_{FalseAlarm}$ from the implausible filler sentences. We applied a liberal criterion and required only that our participants show some discrimination, i.e., a d-prime greater than zero.

2.5. METHODS: PREFERENTIAL LOOKING. In the preferential looking task, the names of the response categories were displayed on screen while participants listened to each sentence. The response labels were outlined in large boxes that matched the color of the corresponding response key. Stimulus presentation and data acquisition were controlled via a custom Python script in PsychoPy (Peirce 2007). Participants were recorded by a webcam embedded in the laptop using Apple Quicktime 7 Pro. Most participants wore acoustically-shielded headphones; participants who preferred not to wear headphones heard the stimuli from a pair of portable desktop speakers.

Videos were aligned to the trial audio using ELAN (Max Planck Institute for Psycholinguistics, Nijmegen; Sloetjes & Wittenburg 2008). Four undergraduate annotators coded the video frame by frame, classifying the direction of participant gaze as: *left, right, up, down, away,* or *can't tell*. Annotators jointly developed a set of coding criteria on the basis of five standard videos. Two to four annotators coded each video. For each frame coded by four annotators, gaze was classified as 'unclear' if fewer than three annotators agreed. For each frame coded by two or three annotators, gaze was classified as 'unclear' if there was any disagreement among annotators.

Our technique departed from the standard practice of displaying images or videos that depict competing referents or situations (in preferential looking: Golinkoff et al. 1987; in the visual-world paradigm: Tanenhaus et al. 1995). We had two reasons for this. The first was practical and related to the challenges of working on a small language like Chamorro. Our construction of linguistic stimuli, which took several months, involved several cycles of elicitation, norming, recording, re-recording, etc. that had to be coordinated among the three team members, who were not always in the same place. To further constrain this process by making the stimuli depictable (and then depicting them) seemed unsustainable. The second reason relates to the nature of wh-dependencies: such constructions are often felicitous only in circumscribed discourse contexts, so a visual scene would have to be carefully calibrated to avoid creating strong extra-linguistic expectations. In the past, such expectations have complicated the interpretation of time-course data on the processing of wh-dependencies (Sussman & Sedivy 2003, Omaki et al. 2010).

We computed the probability of looking in each of the pre-defined gaze locations over 30 ms time windows. For each trial, we aligned the series of windows to the acoustic onset

of the verb. Based on the probabilities in each window, we derived two measures: (i) a *ti måolik* preference score (TMP), defined as the log ratio of looks toward the *ti måolik* ('not good') response category versus looks toward either the *måolik* ('good') response category or center of the screen; and (ii) a *looks-down* preference score (LDP), defined as the log ratio of looks down toward the keyboard versus looks toward the screen (either response category or center of the screen). Both TMP and LDP time series were analyzed with (separate) generalized additive models with several smooth covariate terms: elapsed time from verb onset, the two-term interactions of elapsed time with Argument Plausibility or Extraction Type, and the three-term interaction of elapsed time with Argument Plausibility and Extraction Type. This model structure allowed us to assess whether looks changed over time given the plausibility of the filler and the morphological form of the verb. Models were estimated with the *mgcv* package in R (Wood 2011) using penalized regression splines and parameter selection by generalized cross validation.

Annotators excluded certain videos from analysis, by consensus, because they were largely uncodeable. Uncodeability was attributed to various incidental, ambient aspects of the experimental sessions, such as a participant not paying full attention to the screen. In most of the excluded videos, the lighting was suboptimal; for example, some participants wished to take the experiment on a porch or in direct sunlight. In the end, videos supplied by 45 of the 72 participants were coded. Anonymized data files and the analysis scripts used to generate the tables, charts, and statistical models reported here are available from the authors' website and deposited in their institutional repository.

3. ON-LINE COMPREHENSION OF WH-AGREEMENT: SELF-PACED LISTENING. Forty speakers of Chamorro, ranging in age from 19 to 50, took part in the self-paced listening task. Thirty-one speakers were women. Eleven speakers were born on Rota; 14 on Saipan; 7 on Tinian; 6 on Guam; birthplace information was missing for 2 speakers. Four individuals were excluded from analysis based on the d-prime criterion described in section 2.4.

We analyzed listening times for each of the four segments: the filler (= the wh-phrase), the verb, the adjunct phrase immediately following the verb (XP1), and the disambiguating phrase (XP2). Figure 3 shows listening times at each of these segments for transitive wh-

dependencies, and Table 3 presents the statistical analysis. Figure 4 shows listening times for intransitive wh-dependencies, and Table 4, the statistical analysis. In brief, the results are these. When the verb shows Wh-Agreement, the plausibility of the filler is taken into account immediately after the verb. When the verb could have shown Wh-Agreement but does not, there are no plausibility effects. Finally, when the verb could *not* have shown Wh-Agreement, plausibility effects are present but smaller.

3.1. SELF-PACED LISTENING RESULTS: TRANSITIVE WH-DEPENDENCIES. In wh-dependencies formed from transitive verbs, the verb region showed no effect of Argument Plausibility. (This is the first segment where such an effect could have been observed, since Argument Plausibility is a relation between the filler and the verb.) However, we found a crucial effect of Argument Plausibility in the post-verbal adjunct phrase, XP1, which was identical across all conditions. At this point, a possessor extraction would still be ambiguous from the standpoint of the comprehender, since the filler could still potentially be interpreted as the verb's internal argument. Significantly, participants showed a sensitivity to goodness-of-fit between the filler and the verb just for verbs with overt Wh-Agreement—that is, just for object extractions, not for possessor extractions. This is reflected in the significant simple effects of Argument Plausibility and Extraction Type, as well as their interaction, reported in Table 3 (highlighted with a pointer symbol). We also conducted pairwise comparisons of Argument Plausibility, and found that the difference was significant for verbs with overt Wh-Agreement (INTERCEPT: 7.08 ln ms ± 0.10 ln ms, t = 69, PLAUSIBILITY: .15 ln ms ± .05 ln ms, t = 3.3, p < .005) but not for verbs with no Wh-Agreement (INTERCEPT: 6.93 ln ms ± 0.10 ln ms, t = 71, PLAUSIBILITY: .03 ln ms ± .05 ln ms, t = .85, p = .40).

<INSERT FIGURE 3 HERE>

Why did the effect of Argument Plausibility emerge in the XP1 region but not on the verb itself? We take this to be an effect of the moving-window technique: self-paced listening, like self-paced reading, can be prone to delay effects. This view is supported by the relation of the listening times in our data to word frequency, a lower-level stimulus property. In a separate survey, we obtained subjective frequency ratings from 35 Chamorro

speakers for the verbs in the item sets, in their citation form (see Supplementary Materials). Subjective frequency ratings have been shown to correlate with lexical decision and naming latencies (Gordon 1985, Connine et al. 1990), sometimes more strongly than frequencies obtained from a text corpus. We estimated a simple linear model of duration-residualized RT as dependent on the subjective frequency rating for each of the 24 verbs in the item sets. This relationship was not significant ($p \sim 0.54$). However, on the adjunct phrase immediately following the verb, there was a significant negative correlation between subjective frequency and (duration residualized) listening times: lower frequency verbs were associated with longer listening times ($p < .05$, $R^2 = 0.25$). There was no difference between transitive and intransitive verbs. These results strengthen the case that many effects of verb processing are manifested in listening times not on the segment containing the verb itself, but rather on the immediately following segment.

Although the verb region showed no effect of Argument Plausibility, it did show an effect of Extraction Type: listeners took longer to advance to the next segment when the verb showed overt Wh-Agreement.

<INSERT TABLE 3 HERE>

The fact that verbs with overt Wh-Agreement showed longer listening times than verbs with no Wh-Agreement can be attributed to one or more of the following. First, differences in morphological complexity. In our stimuli, verbs with overt Wh-Agreement show both the infix –*in*- and agreement suffixes, whereas verbs with no Wh-Agreement show only agreement proclitics. Second, a potential difference in the relative frequencies of the two verb forms. Third, a difference in average acoustic duration. In our stimuli, the audio segments containing verbs with overt Wh-Agreement were longer than segments containing verbs with no Wh-Agreement (mean difference: 160 ms, s.d.: 163 ms; paired $t(24)$: 5.0, $p < .001$). We conjecture that the smaller effect of Extraction Type in the filler region is a planning correlate of the longer duration of the verb.

Fortunately, we can show that duration alone cannot explain the difference in listening times between the two verb forms. Participant listening times were regressed against duration using data from the entire experiment, but excluding the target transitive stimuli.

A mixed-effects model of ln(RT) was estimated, with fixed effects of segment duration, and nested effects of Subject and Region, each with a random slope. The (predictive) residuals from that model were then used as the response variable in a subsequent model of just the verb region, dependent now upon the manipulated experimental factors. Note that it would have been misleading to enter stimulus duration directly as a covariate in our original model, since it was correlated with Extraction Type; it was for this reason that we residualized the listening times from the entire experiment. In our model of the residuals, we found that Extraction Type (i.e. verb form) remained a significant predictor (0.08 ± 0.03, $p < .05$). We conclude that the longer duration of verbs with overt Wh-Agreement did not, in and of itself, entirely explain the longer listening times.

In sum, in transitive wh-dependencies, the plausibility of the filler is taken into account immediately after the verb just when the verb shows overt Wh-Agreement. In addition, verbs with overt Wh-Agreement show longer listening times than verbs without it.

3.2. SELF-PACED LISTENING RESULTS: INTRANSITIVE WH-DEPENDENCIES. In wh-dependencies formed from intransitive verbs, we also found an effect of Argument Plausibility in the XP1 region, but the effect was smaller and did not distinguish between extraction types.

Our intransitive stimuli differed from our transitive stimuli in length: intransitive wh-dependencies with subject extraction contained just one adjunct phrase following the verb, whereas transitive wh-dependencies with object extraction contained two (compare 13 with 11).[5] Nonetheless, in the intransitive stimuli there was always one post-verbal adjunct phrase (XP1). At the point when XP1 was encountered, possessor extraction would still be ambiguous from the standpoint of the comprehender: the filler could potentially be interpreted as the subject argument of the verb. Hence we can perform an analysis analogous to the transitives.

The verb region showed no simple or interaction effects. The XP1 region showed a large effect of Extraction Type: subject extractions were read more slowly than possessor extractions. However, because of the stimulus design, these were exactly the extractions in which the XP1 region was the last segment of the utterance. Crucially, there was also a small but significant simple effect of Argument Plausibility in the XP1 region, which was unqualified by an interaction (highlighted in the table with a pointer).

<INSERT FIGURE 4 HERE>

In the intransitive wh-dependencies, in other words, participants showed a sensitivity to goodness-of-fit between the filler and the verb in both subject extractions and possessor extractions. In contrast, in the transitive wh-dependencies, this sensitivity emerged only when the verb showed overt Wh-Agreement—in object extractions but not possessor extractions. On the face of it, this pattern of results suggests that Wh-Agreement affects the comprehension of wh-dependencies, in a way that hastens interpretation of verb-argument combinations or aids disambiguation.

In addition, whether the verb was transitive or intransitive, there were longer listening times in the region immediately following it when the filler was an argument, i.e., a subject or direct object. This second effect is slightly more difficult to interpret, because of the differences described earlier in the number and length of constituents in our stimuli. For intransitive wh-dependencies, the effect most likely reflects the fact that the XP1 constituent was (signaled by the prosody as) utterance-final when the filler was an argument. For transitive wh-dependencies, we have identified at least two causes: first, that there is an effect of Wh-Agreement morphology itself—an effect that survives residualization of segment duration at the verb; and second, that a wh-dependency has been integrated. We consider these ideas in fuller detail in the discussion.

<INSERT TABLE 4 HERE>

A final pattern deserves comment: in the data from intransitive wh-dependencies, there are significant differences between conditions at the filler itself, before the verb is encountered. Inspection of the (untrimmed) means suggests that a single condition—extraction of a plausible argument—was driving the simple effects and marginal interaction in the untrimmed dataset. We considered the possibility that differences in stimulus duration might have been responsible for this: such differences might be spurious or might reflect an incremental effect in production. As in section 3.1, we estimated a model of ln(RT) as predicted by segment duration and region number (excluding the intransitive

conditions). We then predicted the reading times for our critical intransitive stimuli and then modeled the residual RTs (ln ms) with the experimental factors. The effects observed in Table 4 persisted, suggesting that duration alone cannot explain the effect in the initial segment. We are left to conclude that the elevated mean listening time at the filler in subject extractions when the filler was a plausible argument was a spurious effect.

3.3. SELF-PACED LISTENING RESULTS: SUMMARY. The on-line measures from the self-paced listening experiment point to the following conclusions. First, whether the verb was transitive or intransitive, no effect of argument fit was observed on the verb itself. However, verbs with overt Wh-Agreement showed elevated listening times that were independent of stimulus duration alone. This suggests that it is potentially more difficult to process verbs that show overt Wh-Agreement.[6] Given that such verb forms are morphologically complex, this effect is perhaps not surprising.

Second, in transitive wh-dependencies, the potential goodness-of-fit of the filler as the verb's internal argument is reflected in elevated reading times immediately following the verb (in the XP1 region; see Figure 3 and Table 3). Elevated reading times for a potentially implausible argument were observed only when the verb showed overt Wh-Agreement. This is crucial, because when the verb did not show Wh-Agreement, the same region was incrementally ambiguous between an object extraction analysis and a possessor extraction analysis. The comprehender, it seems, forestalls adopting the object extraction analysis in the absence of overt Wh-Agreement, at least to the extent that adopting such an analysis would lead to an observable disruption in processing.

Third and finally, in intransitive wh-dependencies, the potential goodness-of-fit of the filler as the verb's subject argument is registered by a small effect in the XP1 region in the trimmed dataset. The patterns of listening times here were more subtle. The effect did not change depending on extraction type, which is perhaps unsurprising since the form of the verb was constant. The fact that intransitive verbs could generate an anomaly effect is significant for two reasons. First, it demonstrates that verbs with no Wh-Agreement *can* give rise to anomaly effects. Second, it reveals that the input was temporarily compatible with argument extraction even when the wh-dependency was ultimately a possessor extraction. That in turn reveals that in the absence of overt Wh-Agreement, the parser does

not treat argument extraction and possessor extraction as equally good options. As predicted by the Active Filler Strategy, argument extraction is (temporarily) favored. We return to the difference between transitive and intransitive wh-dependencies in greater detail in section 6. It should be emphasized, though, that the anomaly effect was smaller for intransitive wh-dependencies than for transitive wh-dependencies, and only apparent in the log-transformed data. These qualifications suggest that it is less robust than the transitive anomaly effect.

We now turn to data from the preferential looking task.

4. ON-LINE COMPREHENSION OF WH-AGREEMENT: PREFERENTIAL LOOKING. Seventy-two speakers of Chamorro, spanning an age range from 20 to 81, participated in the preferential looking task. Forty speakers were women. Twenty speakers were born on Rota; 46 on Saipan; 2 on Tinian; 3 on Guam; information was missing for 1 speaker. As described in section 2.5, videos of 45 of the 72 participants were codeable for analysis. Four additional participants were excluded based on the d-prime criterion described in section 2.4. Forced choice behavior was analyzed regardless of the codeability of participants' videos (see section 5). Overall, the preferential looking data are more difficult to interpret than the self-paced listening data. Nonetheless, the results from this methodology are consistent with the self-paced listening results.

In brief, participants' looks suggest the following: When the verb shows Wh-Agreement, the plausibility of the filler is taken into account immediately after the verb. When the verb could have shown Wh-Agreement but does not, plausibility effects are not apparent until further downstream, at the offset of XP1. Finally, when the verb could *not* have shown Wh-Agreement, plausibility effects are present from the verb onward.

4.1. PREFERENTIAL LOOKING RESULTS: TRANSITIVE WH-DEPENDENCIES. In analyzing the data, we first considered whether participants' tendency to look toward the *ti måolik* ('not good') response category was sensitive to Argument Plausibility and Extraction Type. We analyzed the average *ti måolik* preference (TMP) score for each condition across the duration of the trial, which is reported in Figure 5. We estimated a generalized additive model of these data, with smooth predictors for elapsed time and the pairwise interactions

of elapsed time with Argument Plausibility, with Extraction Type, and finally the three-way interaction among all three. All predictors were statistically significant, including the three-way interaction ($p < .001$). The three-way interaction is crucial because it tests whether Wh-Agreement morphology affects how contrasts due to Argument Plausibility might unfold over time. We also estimated a model *without* the three-way interaction. Such a model fared worse on all model comparison criteria: for example, the model without the interaction had a higher Akaike Information Criterion score (-324 v. -420; lower is better) and a lower adjusted R^2 (72% v. 79%; higher is better).

Inspection of the Argument Plausibility trendlines in Figure 5 (bottom panel) reveals two different patterns of change over time. When the verb shows overt Wh-Agreement (i.e. object extractions), participants' TMP after the verb onset was actually smaller when the filler was implausible as the verb's argument than when it was plausible. However, beginning with the verb offset, the TMP score steadily increases until the end of the utterance. In contrast, when the verb shows no Wh-Agreement (i.e. possessor extractions), there was no difference in TMP between plausible vs. implausible arguments from the onset of the verb until the offset of the post-verbal adjunct (XP1). Intriguingly, for about 500 ms, there is a sharp rise in the TMP score, which returns to 0 before the end of the utterance.

A difference in TMP between object extractions and possessor extractions was observed at the verb, from its onset, which raises the possibility that the source of interaction reported above might not reflect processing after the verb. We therefore performed a more targeted comparison by restricting our analysis to TMP scores just in the adjunct (XP1) region. We defined this as the 550 ms window beginning from 700 ms post-verb onset to 1250 ms post-verb onset, an interval defined by median acoustic onset and offset of XP1. When considering just the adjunct region, we found a significant baseline difference: overall, object extractions had higher TMPs at the beginning of the analysis window ($p < .001$), but there was no significant difference due to plausibility. There were significant smooth effects only of elapsed time ($p < .001$) and its interaction with Argument Plausibility and Extraction Type ($p < .001$). A model without this interaction fared worse (AIC: -134 v. -154; R^2_{ADJ}: 40% v. 57%). An analysis of just the object extractions revealed a significant interaction between elapsed time and plausibility ($p < .001$). In the

corresponding analysis of possessor extractions, the interaction was not significant ($p = .14$).

The TMP scores reveal a more nuanced picture than the means for self-paced listening. When the verb shows overt Wh-Agreement, the widening contrast between TMP scores for implausible vs. plausible arguments began in the XP1 region, which seems compatible with the anomaly effect in self-paced listening. However, when the verb shows no Wh-Agreement, the fact that there was a brief difference between TMP scores for plausible vs. implausible argument conditions in the possessed DP region does not straightforwardly compare with the lack of anomaly effect in the self-paced listening data. It may thus be necessary to alter the conclusion suggested by the self-paced listening data, namely, that there is *no* effect of anomaly when the verb shows no Wh-Agreement. Before moderating that point, however, we consider one final analysis: the Looks-Down preference scores.

<INSERT FIGURE 5 HERE>

<INSERT FIGURE 6 HERE>

The Looks-Down Preference score (LDP) indexes the trend for participants to look away from the screen down to the keyboard. We calculated this measure when it became clear that participants were increasingly looking away from the screen and to the keyboard as each trial wore on. Figure 6 reports the average LDP across all conditions in the upper-right panel, where it is monotonic increasing across the entire trial. We tentatively attribute this trend to participants' response readiness, their decision-making process, or both. When the filler is (temporarily) implausible, LDP increases much more dramatically when the verb shows overt Wh-Agreement than when it does not (top-right versus bottom-right panels). When the verb shows overt Wh-Agreement (i.e. object extractions), there is a sharp increase in LDP for implausible arguments compared to plausible arguments, beginning in the XP1 region. When the verb shows no Wh-Agreement (i.e. possessor extractions), there was a slight increasing trend across the same interval; but, interestingly, shortly after the possessed DP is introduced, participants quickly shifted their attention predominantly to the screen. Consistent with these descriptions, the best-fitting generalized additive model includes significant smooth terms ($p < .001$) for interactions of

elapsed time with Extraction Type and with Argument Plausibility, and the three-way interaction term. As with the TMP score, the three-way interaction is crucial because it tests whether Wh-Agreement morphology affects how any Argument Plausibility contrasts unfold over time. When we estimated a model *without* the three-way interaction, it fared worse on all model comparison criteria: the model with the interaction had a lower Akaike Information Criterion (-720, v. -525) and a higher adjusted R^2 (82% v. 73%). A targeted analysis of just the XP1 region (analogous to the TMP analysis above) revealed a statistically analogous pattern.

The pattern of LDP over time depended on the presence of overt Wh-Agreement on the verb. When Wh-Agreement was present, participants shifted their attention to the response keys when they heard a wh-dependency with an implausible argument (as opposed to a plausible argument). In contrast, when Wh-Agreement was absent, there were no large plausibility-related differences until after the offset of the possessed DP. Then, participants tended to shift their attention away from the response keys and to the screen. This trend is consistent with the fact that response times were longer for possessor extractions overall (see section 5.1).

4.2 PREFERENTIAL LOOKING RESULTS: INTRANSITIVE WH-DEPENDENCIES. Intransitive wh-dependencies showed increases in the TMP scores for implausible arguments,[7] beginning with the verb and continuing through the post-verbal adjunct. Recall that intransitive verbs do *not* show overt Wh-Agreement, and consequently subject extractions and the corresponding possessor extractions are string-identical through the adjunct region.

Figure 7 depicts the smooth interaction terms of the generalized model considering elapsed time, its interaction with Argument Plausibility, Extraction Type, and the three-way interaction among all three factors. In both subject extractions and possessor extractions, there was a marked tendency to *not* look at the *ti måolik* ('not good') response category during the verb, followed by a corresponding rise to net positive TMP scores. The pattern of TMP scores for intransitive verbs, regardless of extraction type, resembles the pattern for transitive verbs with overt Wh-Agreement: sensitivity to anomaly at least as early as the post-verbal adjunct phrase. In the model of TMP over time, all effects were significant, including the significant interaction between time, Extraction Type, and Argument

Plausibility ($p < .001$). The model without this interaction fared much worse (R^2_{Adj}: 60% v. 70%; AIC: -95 v. -165).

However, inspection of the plots in Figure 7 (and the raw data) reveal that some care is required in interpreting this interaction. In particular, the interpretation is not the same as for transitive wh-dependencies, where there was no positive or increasing TMP score in the XP1 region. For intransitive wh-dependencies, subject extraction appears to have an earlier rise in TMP, confirmed by a model confined to just the verb, and a different time course in the adjunct, also confirmed by a model of just that region. However, it is important to note that the intransitive stimuli were not acoustically as well matched as were the transitive stimuli. In the intransitive stimuli, both the verb and the adjunct had longer durations in possessor extractions than in subject extractions. (The difference between median boundary times was 47 ms for the verb and 286 ms for the adjunct; by contrast, the maximum difference in the transitive stimuli was 16 ms.) This most likely reflects the fact that the subject extractions in our intransitive stimuli were shorter in string length than the possessor extractions—a difference linked to the types of verbs that we selected, on the basis of their ability to host possessor extraction (see note 5). As a final point of similarity, we observe that the peak in TMP scores occurs at approximately the same time for subject extractions and possessor extractions: 900 ms from verb onset. As in self-paced listening, we conclude that comprehenders actively completed the wh-dependency.

<INSERT FIGURE 7 HERE>

5. FORCED CHOICE SENSICALITY RESULTS**.** In sections 3 and 4, we found evidence that the on-line comprehension of wh-dependencies is sensitive to the presence of Wh-Agreement. On the one hand, this is unsurprising, since Wh-Agreement provides a direct morphological cue to the grammatical relation of the gap. On the other hand, it was surprising to discover that when the verb was a transitive verb with no Wh-Agreement, comprehenders showed decreased (or absent) real-time sensitivity to the goodness-of-fit between the filler and the verb. In languages which have no system of Wh-Agreement, comprehenders show a robust, early sensitivity to verb-argument combinations. Further, in Chamorro, when the verb was

intransitive and could not have shown Wh-Agreement, comprehenders *did* show sensitivity to verb-argument combinations. All this leads us to conclude that there is a paradigmatic effect: in contexts in which Wh-Agreement *can* be overt, its absence is viewed as potentially informative.

We now turn to the forced choice data. In this part of the experiment, participants were under no time pressure, so these results present an opportunity to deepen our understanding of the real-time data. For each verb type (transitive or intransitive), we analyze the endorsement rates—that is, the rate at which participants judged a sentence to be sensible (*måolik* 'good')—as well as response times. We then present an analysis of the response patterns based on age of the participant, which we suggest correlates with the content of speakers' grammars.

Below we discuss in greatest detail the results for the preferential looking cohort, which had the largest number of participants and the widest age range. An analysis of the self-paced listening cohort, which led to comparable results, is supplied in the Supplementary Materials.

5.1. FORCED CHOICE SENSICALITY RESULTS: PREFERENTIAL LOOKING COHORT. As mentioned earlier, 72 speakers of Chamorro participated in the preferential looking version of the experiment. Eleven speakers were removed from the forced choice sample because they did not pass the filler criterion described in section 2.4. (The same 11 speakers were also removed from the looking data analysis, although some had already been excluded because of uncodeable video files.)

ENDORSEMENT RATES AND RESPONSE TIMES: TRANSITIVE WH-DEPENDENCIES. The rates at which participants endorsed the transitive wh-dependencies are given in the left-hand columns in the top half of Table 5. Both Extraction Type and Argument Plausibility were significant determinants of participants' endorsement rates ($p < .001$). These two simple effects were qualified by an interaction which reduced the difference between the two Argument Plausibility conditions for possessor extractions ($p < .001$). However, the pairwise difference between levels of Argument Plausibility for possessor extractions did remain significant ($p < .05$; for a model including only Possessor Extraction conditions).

<INSERT TABLE 5>

The response times are presented as median times in the right-hand columns of Table 5. For response times for the transitive wh-dependencies, there was a significant effect of Extraction Type ($p < .005$; model estimated on (natural) log response times), a 2-way interaction of Response Choice and Argument Plausibility ($p < .001$), and a 3-way interaction of Argument Plausibility, Extraction Type and Response Choice ($p < .01$). There was a marginal effect of Argument Plausibility ($p < .10$). This pattern of statistics reflects two generalizations. Participants were faster at giving the expected response; however, this effect was neutralized for possessor extractions when the filler was an implausible argument. In just the possessor extractions with implausible arguments, a pairwise comparison between *ti måolik* ('not good') and *måolik* ('good') responses revealed no statistically significant difference in log RT ($t = 0.13$).

ENDORSEMENT RATES AND RESPONSE TIMES: INTRANSITIVE WH-DEPENDENCIES. The rates at which participants endorsed the intransitive wh-dependencies are given in the left-hand columns in the bottom half of Table 5. Both Extraction Type and Argument Plausibility were significant determinants of participants' endorsement rates (ps $< .001$). These two main effects were qualified by an interaction which reduced the difference between the two Argument Plausibility conditions just in case the filler was a possessor ($p < .001$). In contrast to the transitive wh-dependencies, the pairwise difference between levels of Argument Plausibility for possessor extractions was not significant ($z = 1.30$).

In the response times, there was one reliable effect: an interaction between Argument Plausibility and Response Choice ($p < .01$). There was also a marginal 3-way interaction ($p < .10$). This pattern reflects one major generalization: participants were faster at giving the expected response. There was also a tendency for the difference in response times between the two response choices to be attenuated in possessor extractions when the filler was an implausible argument. While a 3-way interaction coefficient did not achieve significance at $\alpha = 0.05$, a simple pairwise comparison revealed no statistically significant difference between response types in possessor extractions ($t = 0.09$).

5.2. GENERATIONAL DIFFERENCES IN POSSESSOR EXTRACTION. The lower endorsement rates for (even plausible cases of) possessor extraction led us to consider whether all speakers recognized possessor extraction as a construction of Chamorro. In the Supplemental Materials, we present an analysis of individual variation in speakers' endorsement rates of possessor extractions and argument extractions, when those extractions were plausible. In brief, we found no relationship between age and performance on argument extractions, but a significant positive correlation between age and accuracy on possessor extraction. Only the older speakers in our sample, those over 55, showed comparable accuracy on plausible possessor extractions and plausible argument extractions. With this result in hand, we reanalyzed the possessor extractions in which the filler was implausible as an argument of the verb just for those speakers aged 55 and older. These temporarily implausible, globally plausible extractions received lower endorsement rates than the extractions which were both temporarily *and* globally plausible. Endorsement rates and response times for these participants are given in Table 6.

<INSERT TABLE 6>

The general pattern of results is similar to that for the overall population. Plausible possessor endorsement rates increased from 66% to 84%—unsurprisingly, given that this figure was used in the age-based analysis. Crucially, the endorsement rate for Implausible Argument Possessor Extractions remained significantly lower than the endorsement rate for Plausible Argument Possessor Extractions: 61%, versus 84%. In other words, once we controlled for potential differences in speakers' perception of uncontroversially plausible possessor extractions, a difference remained in how they treated the temporarily implausible ones. The full analysis is detailed in the Supplemental Materials.

What do we learn from the analysis of the older cohort's endorsement rates and response times? Recall that the pattern of response choices in the entire population was that of a 'temporary plausibility' effect when the verb was transitive: participants were temporarily sensitive to the plausibility of the filler as the internal argument of the verb, even though the filler was eventually linked to a possessor. No such temporary plausibility effect was found when the verb was intransitive. The analysis of the older cohort

strengthens this interpretation of the pattern. Because older participants showed the temporary plausibility effect, as well as its interaction with transitivity, *and* endorsed plausible arguments at comparable rates for both argument and possessor extractions, it seems unlikely that the temporary plausibility effect in this group stems from incomplete knowledge of possessor extraction.

Finally, recall that self-paced listening participants made the same forced choice sensicality judgment as did the preferential looking participants—though the self-paced listening participants were generally younger. Nonetheless, their pattern of choices was similar. The highest accuracy was observed in argument extractions. As for possessor extractions, when the verb was transitive, there was a difference between whether the filler was temporarily plausible or implausible as the verb's internal argument, despite its being always globally plausible as a possessor. But as in the preferential looking data, when the verb was intransitive, possessor extractions showed no sensitivity to temporary plausibility. A full analysis of the self-paced listening participants' endorsement rates and RTs is given in the Supplemental Materials.

5.3. SUMMARY OF FORCED CHOICE DATA. The outcome of the forced choice task can be summarized as follows:

(a) When the gap was a direct object and the verb showed overt Wh-Agreement, participants were able to successfully classify the filler on the basis of its plausibility as the verb's internal argument. This is reflected both in the endorsements and the response times.

(b) When the gap was the possessor of the direct object and the verb showed no Wh-Agreement, participants' classification was sensitive not only to the filler's plausibility as a possessor (always globally plausible) but also to its plausibility as the verb's internal argument. This is reflected both in the non-parity of the endorsement rates and in the equivocal response times for the response choicess.

(c) When the gap was the subject or the possessor of the subject, participants' classification was sensitive only to the filler's plausibility as interpreted at the actual gap site.

(d) Endorsement rates for possessor extraction were lower overall than endorsement rates for argument extraction.

The off-line measures of forced choice RT and accuracy present an interesting contrast to the on-line data of per-segment listening times or looking preference. Notably, the forced choice data suggest that implausible arguments were processed differently than plausible arguments when the verb showed no Wh-Agreement. However it is difficult to pinpoint such a difference in the on-line data. This suggests either that some incremental process distinguished plausible and implausible arguments but we could not detect it with our design or methods; or else that some later process, perhaps at the offset of the utterance, distinguished the two conditions. We take up these issues below.

6. DISCUSSION. Wh-dependencies are characteristically affiliated with prolonged ambiguity because they involve the syntactic separation of the filler from the gap. Previous research has indicated that comprehenders negotiate this ambiguity by engaging in the eager formation and testing of hypotheses about the gap site. This strategy, identified as the Active Filler Strategy (Frazier 1987), is guided by a limited set of cues sourced from the grammar and the lexicon. In this study we tracked the real-time comprehension of wh-dependencies in the Austronesian language Chamorro. We wanted to see how Chamorro grammar might contribute to our understanding of this process, because it provides different kinds of cues to the gap site than do previously studied languages. Specifically, Chamorro has an elaborate system of Wh-Agreement which registers both the presence of a wh-dependency and the grammatical relation of the gap. By virtue of this morphological information, Wh-Agreement provides early, reliable cues to the resolution of wh-dependencies. In contrast to most cues studied to date, Wh-Agreement morphology can potentially be apprehended before lexical access to the verb stem. Finally, in object extraction, Wh-Agreement morphology can be omitted without truth-conditional consequences. This allows for a very precise assessment of its contribution to the comprehension of wh-dependencies. In brief, we found that morphological information does have temporal priority over syntactic information, such as word order, and semantic information, such as argument structure, in the unfolding parse.

We tested for the effect of Wh-Agreement by creating constituent questions in which a verb was paired with a potential filler that was either sensible or anomalous as its argument (Garnsey et al. 1989, Traxler & Pickering 1996, Wagers & Phillips 2009). We measured the contrast between those pairings in two experiments: a self-paced listening experiment, which gives rise to a phrase-by-phrase measure of processing complexity, and a preferential looking experiment, which more directly tracks perceived anomaly. The results across the two measures were consistent, and they were surprising.

First, there was a strong, immediate contrast between the sensible and anomalous verb-filler combinations when the verb was overtly inflected for Wh-Agreement. In self-paced listening, listening times at the post-verbal adjunct phrase were elevated when the filler was an implausible argument. In preferential looking, a preference to look to the *ti måolik* ('not good') category increased from the offset of the verb until the offset of the utterance. This effect was amplified by a dramatic increase in looks to the response keys themselves. The self-placed listening effect is analogous to the anomaly effects routinely observed in reading studies of wh-dependencies, whether the measure is reading times (Traxler & Pickering 1996, Wagers & Phillips 2009) or evoked potentials (Garnsey et al. 1989). Taking the self-paced listening and preferential looking results together, we see that the anomaly caused by the implausible argument was detected relatively early with respect to the verb. More important, the anomaly was detected well before the offset of the utterance, before conclusive bottom-up evidence of the gap site could be encountered. In short, when confronted with a wh-dependency with overt Wh-Agreement, Chamorro comprehenders show performance patterns like those shown by speakers of well-tested languages (cf. Phillips & Wagers 2007).

Second, Chamorro speakers behave quite differently when processing transitive verbs which do not show overt Wh-Agreement. These verb forms show normal subject-verb agreement, but do not positively indicate the presence of a gap. In a wh-dependency, these verb forms provide the same type of information about the gap as would be provided by an overt subject in better-studied SVO languages: namely, that the gap is not the subject of that verb. On the verb itself, there was no anomaly effect, in either self-paced listening or preferential looking. Nor was there an anomaly effect on the post-verbal adjunct phrase. In preferential looking, there was a brief anomaly effect in the possessed DP region, in terms

of a reliable increase in *ti måolik* ('not good') preference scores departing from a steady baseline but not in terms of the *Looks Down* preference scores.

There was a third pattern when speakers processed intransitive verbs, which (for subject and possessor extractions) *could not* show Wh-Agreement. For these verbs, there were weak but consistent anomaly effects on the post-verbal adjunct phrase in both self-paced listening and preferential looking. Importantly, these effects were uniform across subject extractions and possessor extractions. There were thus contrasts in the timing, duration, and kind of responses that comprehenders displayed to verbs with overt Wh-Agreement as compared to verbs without it. As the intransitive conditions show, there was also a distinction between verbs which could have shown Wh-Agreement and verbs which would never have shown Wh-Agreement. These data suggest that the Active Filler Strategy is sensitive to the organization of the Wh-Agreement paradigm. Specifically, when a transitive verb could have shown Wh-Agreement, but does not, its failure to do so constitutes a source of evidence that the gap is not a direct object.

Table 7 gives a visual summary of these findings across all measures.

<INSERT TABLE 7 HERE>

6.1. CUES PROVIDED BY THE WH-AGREEMENT PARADIGM. We now consider the role of Wh-Agreement itself in dependency formation.

Our results suggest that overt Wh-Agreement *facilitates* the processing of wh-dependencies. It might do so in several ways. The morphology could signal that there is an open wh-dependency that needs resolution. It could serve to resolve the ambiguity of the gap site, because Wh-Agreement encodes the grammatical relation of the gap. Finally, it could instantiate an open structural licensing requirement (in the spirit of Berwick et al.'s (1991) principle-based parsing, or Merlo and Stevenson's (2000) constraint-based architecture). The idea behind this last possibility is that the presence of a marked feature, like Wh-Agreement morphology, requires an appropriate grammatical licensing context and establishing this context becomes a parser priority. This could be viewed as a generalized version of the Active Filler Strategy (Frazier & Flores D'Arcais 1989). Specifically, we hypothesize that overt Wh-Agreement morphology leads to the projection

of the VP-internal structure that would license the agreement—i.e.,. an object gap that can allow the filler to be linked to the verb's internal argument.

Our results further suggest that comprehenders are affected by the absence of Wh-Agreement morphology when it could have been overt. Object extraction only optionally gives rise to overt Wh-Agreement, and possessor extraction never gives rise to it. Given this ambiguity in the *absence* of Wh-Agreement, the comprehender may be uncertain about the right analysis. In other words, the mere possibility of possessor extraction may be sufficient to forestall an object analysis. This explanation would align with recent psycholinguistic theories that link complexity and predictability (Hale 2001, 2006; Levy 2008). The key insight of these theories is that the difficulty of incorporating a new word into the sentence depends on how that new word affects the existing distribution of parses. Given only a filler, such as *håyi na påtgun* 'which child', there are a range of possible continuations, weighted by their likelihood or by some other function. Imagine that these continuations are categorized according to the grammatical relation of the gap. A transitive verb with no Wh-Agreement, such as *un låksi* 'you sew', removes the possibility of a subject gap and an oblique gap. A verb inflected for object Wh-Agreement, such as *linaksem-mu,* removes not only these possibilities but also the possibility of a possessor gap. The role played by Wh-Agreement in limiting continuations thus provides an attractive account of how it is processed on-line. While predictability-based theories have been mute about the link between the distribution of parses and interpretations, we suppose that comprehenders compute interpretations only for parses that are sufficiently highly weighted.

Our data thus admit two complementary explanations of how the Wh-Agreement paradigm could affect on-line processing of unresolved wh-dependencies: overt Wh-Agreement could promote a particular analysis, or its absence could restrain an analysis. It is probably not possible to tease apart these two explanations in our current dataset. Some challenges should be highlighted for the predictability theory, however. First, intransitive verbs did give rise to an anomaly effect, even though intransitives are generally compatible with a possessor continuation. We might attribute to intransitive verbs, therefore, a 'baseline' active filler effect associated with identifying the subject. Second, while we know that a verb with overt Wh-Agreement reduces the odds of possessor extraction to nil, we do not know exactly what the impact is of a normally inflected verb on the odds of possessor

extraction. If we reason by Bayes' Rule, the posterior odds in favor of possessor extraction are equal to the product of two terms: (i) the prior odds in favor of possessor extraction, calculated before the verb is considered; and (ii) the likelihood ratio, i.e., the ratio between the conditional probability of the verb given a possessor extraction and the conditional probability of the verb given an object extraction. In the absence of a Chamorro corpus large enough for us to determine any of these ratios, we must constrain our beliefs in other ways. We attempted to estimate the likelihood ratio by conducting a preference survey in which Chamorro speakers were asked to choose between two truth-conditionally equivalent sentences. Thirteen speakers completed a survey with 12 pairs of items: 3 pairs contrasted object extraction with or without overt Wh-Agreement, 3 pairs contrasted extraction of the possessor of the direct object with the periphrastic form illustrated in 17b; and 6 pairs contrasted subjects that are immediately post-verbal with subjects that occur clause-finally.

(17) **Sample pair of items from survey: two ways to question a possessor**

a. Håyi na famalåo'an un kastiga [famagu'un-ñiha _] gi iskuela?
who? L women AGR punish children-AGR LOC school
'Which women did you punish children of at school?'

b. Håyi na famalåo'an manggai [famagu'un un kastiga gi iskuela] _ ?
who? L woman AGR-have children AGR punish LOC school
'Which women have children who you punished at school?'

Among the object extractions, the form with no Wh-Agreement was preferred 72% of the time. Among the two ways to question the possessor, the form illustrated in 17a was preferred 74% of the time. If these quantities can serve as estimates of the conditional probability of a form, then the likelihood ratio turns out to be essentially one (.74/.72). In other words, if a verb does not show Wh-Agreement, then the odds in favor of a possessor extraction analysis are probably close to the simple prior likelihood of possessor extraction: which is to say, probably quite small. Note that the conclusion might turn out otherwise if we had a better estimate: if speakers overwhelmingly preferred overt Wh-Agreement for object extraction, then the *absence* of Wh-Agreement would be much more informative. In other words, the degree to which the parser reserves commitment to the object analysis depends on the relative abundance of possessor extractions compared to

object extractions. We suspect that the prior likelihood is strongly *against* possessor extraction in general. If possessor extraction were so rare that the posterior odds of a possessor analysis were essentially nil, a predictability-based account of our data would not be that compelling. But much remains to be investigated here.

6.2. POSSESSOR EXTRACTION. On-line measures suggested that transitive verbs with no Wh-Agreement did not give rise to as robust an anomaly effect as verbs with overt Wh-Agreement. However, off-line sensicality judgments clearly reflected a sensitivity to the goodness-of-fit between the filler and the verb. We now consider why that should be the case.

Our first explanation is what we call *imperfect analysis*. In this theory, when comprehenders encounter a transitive verb with no Wh-Agreement, they very seldom make commitments to an analysis beyond the input at the verb. That is, they do not pursue an object interpretation of the filler. Once they encounter the sentence-final possessed DP, which serves as the verb's object, they must link the filler to this DP as its possessor. We conjecture that on some substantial fraction of trials, they fail at doing this and do not impose an exhaustive grammatical analysis on the string. In such trials, their off-line response would be sensitive to goodness of fit, causing them to endorse plausible arguments at a higher rate than implausible arguments.

Our second explanation we call *failed reanalysis.* In this theory, when comprehenders encounter a transitive verb with no Wh-Agreement, they do construct the syntactic representation of an object extraction, but they do not pursue the corresponding semantic interpretation because of remaining uncertainty. They sometimes later fail to reanalyze to the possessor analysis when confronted with that evidence. As a consequence, their off-line response is based on the syntactic object analysis that they initially constructed.

Both species of explanation recognize that comprehenders face difficulty with the possessor analysis in transitive wh-dependencies. But could either explanation square with the further observation that, in intransitive wh-dependencies, there was no interaction between extraction and plausibility? There are two reasons to think that the answer is yes. For the failed reanalysis account, we conjecture that reanalysis is more difficult when the comprehender has analyzed the filler as the object of a transitive verb, a complement, but

must reanalyze it as the possessor, a specifier. In contrast, it is made easier when the comprehender has analyzed the filler as the subject of an intransitive verb, a specifier, and must then reanalyze it as a possessor, also a specifier. We think of the first, more difficult reanalysis as analogous to the well-studied NP/Z and NP/S ambiguities (Sturt et al. 1999). In the NP/Z ambiguity, an NP is initially attached as the object of the verb inside a sentence-initial adjunct clause (*While the man darned the socks …*), but then must be reanalyzed as the matrix subject (*While the man darned the socks fell off his lap*). Such sentences are notorious garden paths. In an NP/S ambiguity, an NP is initially attached as the object of the verb that can also take a complement clause (*Jim forgot his suitcase …*); but then must be reanalyzed as the embedded subject (*Jim forgot his suitcase was stored in the closet*).

Whether the imperfect analysis or failed reanalysis is the correct account may have some bearing on debates on 'good enough' processing (Christianson et al. 2001). Both NP/Z and NP/S garden paths lead to lingering misinterpretations (Christianson et al. 2001, Sturt 2007). Recent evidence from English suggests that this is because reanalysis does not expunge the truth-conditional commitments of the initial syntactic representation (Slattery et al. 2013)—which would be most consistent with our failed reanalysis explanation. For either explanation, the relatively low frequency of extraction of possessors of objects could help explain why comprehenders sometimes fail to achieve the correct analysis. Extraction of possessors of intransitive subjects, on the other hand, is considerably more frequent in attested examples (see, for instance, the Chamorro Dictionary Database). We leave this issue as an open question.

6.3. CONCLUSION. Experimental evidence from two on-line sentence processing tasks suggests that the morphological paradigm of Wh-Agreement affects the speed and accuracy with which Chamorro wh-dependencies are recovered. There were significant time-course differences both when the morphology was overt and when it was not. These differences have consequences for theories of real-time grammatical licensing and syntactic prediction, like the Active Filler Strategy, and how these interact with semantic interpretation. In addition, by involving a demographically broad section of the community in our off-line

task, we discovered that many older Chamorro speakers have difficulty understanding extraction of possessors of objects, and many younger speakers do not accept it at all.

More generally, we hope to have shown some of the positive consequences for the theory of sentence processing when its evidentiary base is broadened to include languages with distinct combinations of typological features. At the same time, traditionally resource-intensive experimental methods can contribute positively to syntactic theory when they are cooperatively situated in, and adapted to, smaller communities of non-Western speakers.

REFERENCES

ANAND, PRANAV; SANDRA CHUNG; and MATTHEW WAGERS. Widening the net: Challenges for gathering linguistic data in the digital age. Response to NSF SBE 2020: Future Research in the Social, Behavioral & Economic Sciences. Online: http://www.nsf.gov/sbe/sbe_2020/.

AOSHIMA, SACHIKO; COLIN PHILLIPS; and AMY WEINBERG. 2004. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51.23-54.

BAAYEN, R. HARALD; and PETAR MILIN. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3.2.12-28.

BATES, DOUGLAS; MARTIN MAECHLER; AND BEN BOLKER. 2012. Lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. Online: http://CRAN.R-project.org/package=lme4.

BOERSMA, PAUL; DAVID WEENINK. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.3.74. Online: http://www.praat.org.

BOLAND, JULIE E.; MICHAEL K. TANENHAUS; and SUSAN M. GARNSEY. 1995. Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language* 34.774-806.

CHATER, NICK; MATTHEW CROCKER; and MARTIN PICKERING. 1998. The rational analysis of inquiry : The case of parsing. *Rational models of cognition*, ed. by Nick Chater, Matthew Crocker, and Martin J. Pickering, 441-468. Oxford: Oxford University Press.

CHRISTIANSON, KIEL. 2002. Sentence processing in a "nonconfigurational" language. East Lansing, MI: Michigan State University dissertation.

CHRISTIANSON, KIEL; and HEE YOUN CHO. 2009. Interpreting null pronouns (pro) in isolated sentences. *Cognition* 112.175-180.

CHRISTIANSON, KIEL; and FERNANDA FERREIRA. 2005. Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98.105-135.

CHRISTIANSON, KIEL; ANDREW HOLLINGWORTH; JOHN F. HALLIWELL; AND FERNANDA FERREIRA. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 4.368-407.

CHUNG, SANDRA. 1982. Unbounded dependencies in Chamorro grammar. *Linguistic Inquiry* 13.39-77.

CHUNG, SANDRA. 1994. Wh-Agreement and "referentiality" in Chamorro. *Linguistic Inquiry* 25.1–44.

CHUNG, SANDRA. 1998. *The design of agreement: Evidence from Chamorro*. Chicago: University of Chicago Press.

CLAHSEN, HARALD; and SAM FEATHERSTON. 1999. Antecedent priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research* 28.415-437.

CLEMENS, LAUREN EBY; JESSICA COON; PEDRO MATEO PEDRO; ADAM MILTON MORGAN; MARIA POLINSKY; GABRIELLE TANDET; and MATT WAGERS. 2014. Ergativity and the complexity of extraction a view from Mayan. *Natural Language and Linguistic Theory,* to appear.

CRAIN, STEPHEN; and JANET D. FODOR. 1985. How can grammars help parsers. *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives*, ed. by D. Dowty, L. Karttunen, and A. M. Zwicky, 94–128. Cambridge: Cambridge University Press.

FERREIRA, FERNANDA.; JOHN M. HENDERSON; MICHAEL D. ANES; PHILLIP A. WEEKS; DAVID K. MCFARLANE. 1996. Effects of lexical frequency and syntactic complexity in spoken language comprehension: Evidence from the auditory moving window technique. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22.324-335.

FIEBACH, CHRISTIAN J.; MATTHIAS SCHLESEWSKY; ANGELA D. FRIEDERICI. 2001. Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research* 30.321–338.

FODOR, JANET DEAN. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9.427-473.

FRAZIER, LYN. 1987. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5.519–560.

FRAZIER, LYN; and GIOVANNI B. FLORES D'ARCAIS. 1989. Filler–driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28.331–44.

GARNSEY, SUSAN M.; MICHAEL K. TANENHAUS; and RICHARD M. CHAPMAN. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18.51-60.

GAGLIARDI, ANNE; AND JEFFREY LIDZ. 2013. Statistical insensitivity in the acquisition of Tsez noun classes. Ms.

GIBSON, EDWARD D. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain,* ed. by Alec P. Marantz, Yasushi Miyashita, and Wayne O'Neil*,* 95-126. Cambridge, MA: MIT Press.

GOLINKOFF, ROBERTA. M.; KATHRYN HIRSH-PASEK; KATHLEEN M. CAULEY; and LAURA GORDON. 1987. The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* 14.23-45.

HARRIS, ALICE C.; and ARTHUR G. SAMUEL. 2011. Perception of exuberant exponence in Batsbi: Functional or incidental? *Language* 87.447-469.

HAWKINS, J. A. 2007. Processing typology and why psychologists need to know about it. *New Ideas in Psychology*. 25.87–107.

HENRICH, JOSEPH; STEVEN J. HEINE; and ARA NORENZAYAN. 2010. *Behavioral and Brain Sciences* 33.61-83.

JAEGER, T. FLORIAN.; and ELISABETH NORCLIFFE. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3.866-887.

LEWIS, M. PAUL; GARY F. SIMONS; and CHARLES D. FENNIG. 2013. *Ethnologue: Languages of the world*, 17th edition. Dallas, Texas: SIL International. Online: http://www.ethnologue.com.

MACMILLAN, NEIL A.; and C. DOUGLAS CREELMAN. 1991. *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

MCELREE, BRIAN; STEPHANI FORAKER; and LISBETH DYER. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48.67-91.

MERLO, PAOLA; and SUZANNE STEVENSON. 2000. Lexical syntax and parsing architecture. *Architectures and Mechanisms for Language Processing,* ed. by M. Crocker & M. Pickering, 161-188. Cambridge, UK: Cambridge University Press.

MUCKEL, SANDRA; and THOMAS PECHMANN. 2002. Predictive antecedent reactivation: Prosody helps to process German verb-final sentences. Ms., Leipzig.

NAKANO, YOKO; CLAUDIA FELSER; and HARALD CLAHSEN. 2002. Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research* 31.531-571.

OMAKI, AKIRA; ELLEN LAU; IMOGEN DAVIDSON WHITE; MYLES DAKAN; and COLIN PHILLIPS. 2010. Hyper-active gap filling: Pre-verbal object gap creation in English filler-gap dependency processing. Ms.

PEIRCE, JOHNATHAN W. 2007. PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods* 162.8-13.

PHILLIPS, COLIN. 2006. The real-time status of island phenomena. *Language* 82.795-823.

PHILLIPS, COLIN; and MATTHEW WAGERS. 2007. Relating structure and time in linguistics and psycholinguistics. *Oxford Handbook of Psycholinguistics*, ed. by G. Gaskell, 739-756. Oxford: Oxford University Press.

PICKERING, MARTIN J.; and MATTHEW J. TRAXLER. 2001. Strategies for processing unbounded dependencies: Lexical information and verb-argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27.1401-1410.

PRITCHETT, BRADLEY. 1992. *Grammatical competence and parsing performance.* Chicago: University of Chicago Press.

R CORE TEAM. 2013. R: Language and Environment for Statistical Computing. Vienn, Austria: R Foundation for Statistical Computing. Online: http://www.r-project.org.

ROHDE, DOUG. 2003. Linger: a flexible platform for language processing experiments, version 2.94. Online: http://tedlab.mit.edu/~dr/Linger/

ROSENTHAL, ROBERT; and RALPH L. ROSNOW. 2009. *Artifacts in behavioral research*. (Reissue of three books: 1. *Artifact in behavioral research*; 2. *Experimenter effects in behavioral research*; 3. *The volunteer subject*). New York: Oxford University Press.

ROSS, JOHN ROBERT. 1967. *Constraints on variables in syntax*. Cambridge MA: MIT dissertation.

SALVERDA, ANNE PIER; DELPHINE DAHAN; and JAMES M. MCQUEEN. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90.51-89.

SEDIVY, JULIE C.; CRAIG CHAMBERS; MICHAEL TANENHAUS; and GREGORY N. CARLSON. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71.109-147.

SLATTERY, TIMOTHY J; PATRICK STURT; KIEL CHRISTIANSON; MASAYA YOSHIDA; AND FERNANDA FERREIRA. 2013. Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*. 69.104-120.

STURT, PATRICK; MARTIN PICKERING; AND MATTHEW CROCKER. 1999. Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*. 40.136-150.

SLOETJES, HAN; and PETER WITTENBURG. 2008. Annotation by category – ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation*.

STOWE, LAURIE A. 1986. Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1.227-245.

STURT, PATRICK. 2007. Semantic re-interpretation and garden path recovery. *Cognition.* 105.477-488.

STURT, PATRICK; MARTIN J. PICKERING; MATTHEW W. CROCKER. 1999. Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language* 40.136–150.

TRAXLER, MATTHEW J.; and MARTIN J. PICKERING. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35.454–475.

WAGERS, MATTHEW. 2008. *The structure of memory meets memory for structure in linguistic cognition*. College Park, MD: University of Maryland dissertation.

WAGERS, MATTHEW; and COLIN PHILLIPS. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics.* 45.395-433.

WAGERS, MATTHEW W.; ELLEN F. LAU; and COLIN PHILLIPS. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61.206-237.

WANNER, ERIC; and MICHAEL MARATSOS. 1978. An ATN approach to comprehension. *Linguistic Theory and Psychological Reality.* ed. by Morris Halle, Joan Bresnan, and George A. Miller, 119–161. Cambridge, MA: MIT Press.

WEINBERG, AMY. 1993. Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research* 22.339-364.

WOOD, SIMON N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametic generalized linear models. *Journal of the Royal Statistical Society B.* 73.3-36.

---

* We are indebted to many people in the CNMI who helped and encouraged us in our research. Heartfelt thanks to Dr. Elizabeth D. Rechebei and the Chamorro Dictionary Working Group, the Chamorro/Carolinian Language Policy Commission, the CNMI Parole Board, and the CNMI Arts Council for their continuing support. On Saipan, we also thank the staff of the Saipan Chamorro Cultural Village Center, Inetnun Åmut yan Kutturan Natibu, the Chamorro bilingual teachers of the Public School System, the Joeten-Kiyu Library, and the residents of As Teo village; on Rota, Tita A. Hocog, the Office of the Mayor, the Chamorro bilingual teachers of Sinapalo School, the staff of the Luta Chamorro Cultural Village Center, and the Public Health Center; and on Tinian, Florine M. Hofschneider and the Tinian Public Library. At UCSC, we are grateful to Pranav Anand and Jim McCloskey, as well as to Sylvia Soule, Sean Hayes, Travis Heller, Shawna Mattison, Mary Moretti, and Natalie Warn. We thank the audiences of presentations at CUNY (2012), GLOW (2012), AFLA (2013), and in the CNMI for their suggestions; and Kiel Christianson and an anonymous reviewer for helpful critical feedback. The research reported here was supported in part by the Hellman Foundation (MW), by the Gary D. Licker Memorial Chair of Cowell College (SC), and by National Science Foundation under Grants No. BCS-0753240 to the University of California, Santa Cruz, and No. BCS-0753594 to the Northern Mariana Islands Council of the Humanities. Supplementary Materials can be accessed at this URL: http://people.ucsc.edu/~mwagers/.

[1] Glosses, orthography, and other symbols. AGR = person/number agreement; D = determiner; L = linker; LOC = local case; POSS = possessor; POSS.AGR = possessor agreement; PROG = progressive; SUBJ.AGR = subject-verb agreement, WH[SBJ/OBJ/OBL] = Wh-Agreement with subject, object or oblique; UNM = unmarked case. <å> is the low back vowel. In wh-dependencies, the extraction site is marked with an underscore (_).

[2] Note that when a verb is nominalized, its subject-verb agreement is chosen from the paradigm for possessor-noun agreement (e.g. –*mu* '2 sg.').

[3] Possessors can be extracted in Chamorro only when the entire possessed DP is a direct object or an intransitive subject, and the D of this possessed DP is the null indefinite article (see Chung 1998).

[4] Mixed-effects models with maximal random effects failed to converge. See Supplementary Materials for alternative analyses.

[5] In Chamorro, the intransitive verbs that allow the possessor of their subject to be extracted are stative (e.g. *maguf* 'happy', *mapotgi'* 'pregnant', *måsa* 'ripe', but not *chålik* 'laugh', *chotchu* 'eat (something)', or *kuentus* 'speak'). In constructing the materials, we found it difficult to get these verbs to co-occur naturally with more than one adjunct phrase. Our experimental design required us to use these intransitive verbs not only in possessor extractions but also in subject extractions. In the end, this meant that our examples of subject extraction contained just one adjunct phrase following the verb, whereas sentences involving object extraction contained two adjunct phrases following the verb.

[6] Another conceivable possibility is that the elevated listening times could reflect a potential incremental ambiguity in Chamorro between verbs with object Wh-Agreement and certain passive verb forms. For example, the object Wh-Agreement form of *guaiya* 'love' is *guinaiyam-mu* (here with a second person subject), which begins in the same way as the passive verb form *guinaiya*: the two forms have the initial syllable [gwi]. However, the two forms are disambiguated in the second syllable by the presence vs. absence of primary stress—an important cue in Chamorro, a language in which most phonological alternations are stress-sensitive. Given that stress can affect lexical disambiguation in speech processing (Salverda et al. 2003), the two forms could, in principle, be disambiguated by the third syllable, before either the acoustic offset of the word or the syllables that would likely identify the verb stem were reached. Importantly, whether or not the comprehender quickly disambiguates to the Wh-Agreement form from a

hypothetical passive competitor, the filler would be identified as the verb's internal argument—which is what matters from the standpoint of argument plausibility.

[7] We do not report the Looks-Down Preference scores for intransitive wh-dependencies, because there were substantial baseline issues for the possessor extractions that rendered that measure difficult to interpret. Data and statistical analysis of this measure are available in the Supplementary Materials.

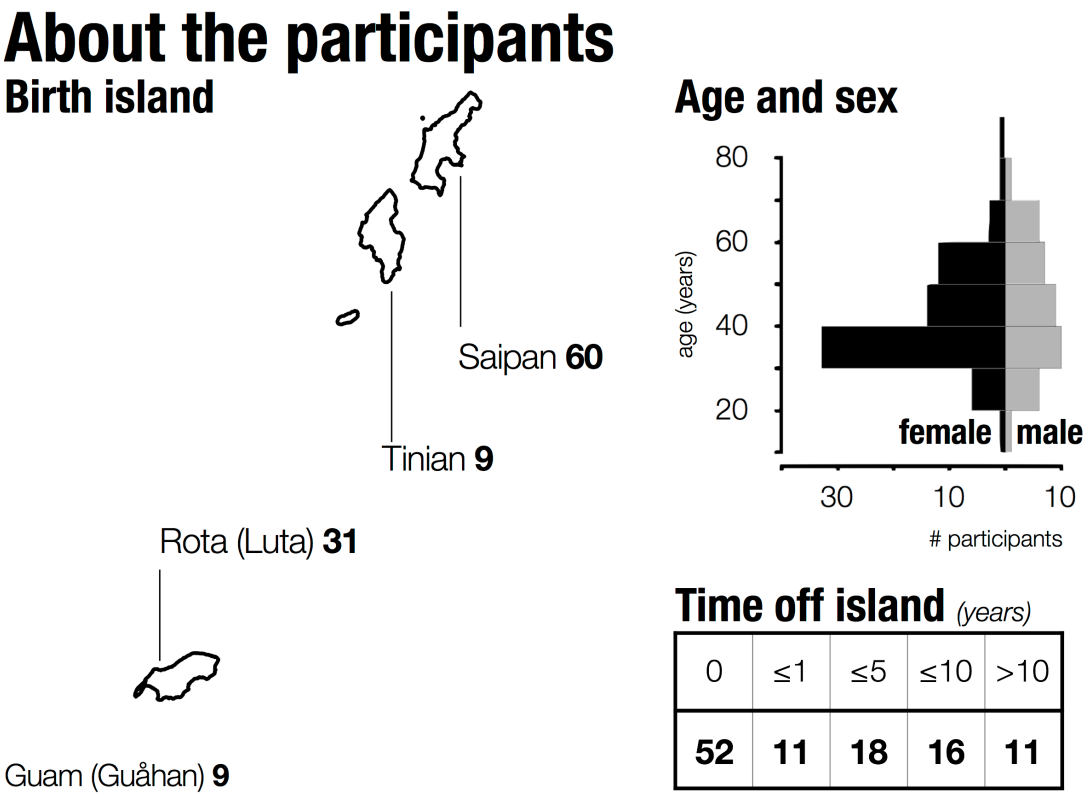

FIGURE 1. Participant demographics.

The map on the left indicates the breakdown of participants by birthplace (Guam not depicted). Charts on the right indicate the breakdown of participants by sex and age (top) and how much time, in years, participants spent outside the Marianas (bottom). Birthplace information was unavailable for three individuals; off-island information was unavailable for four.

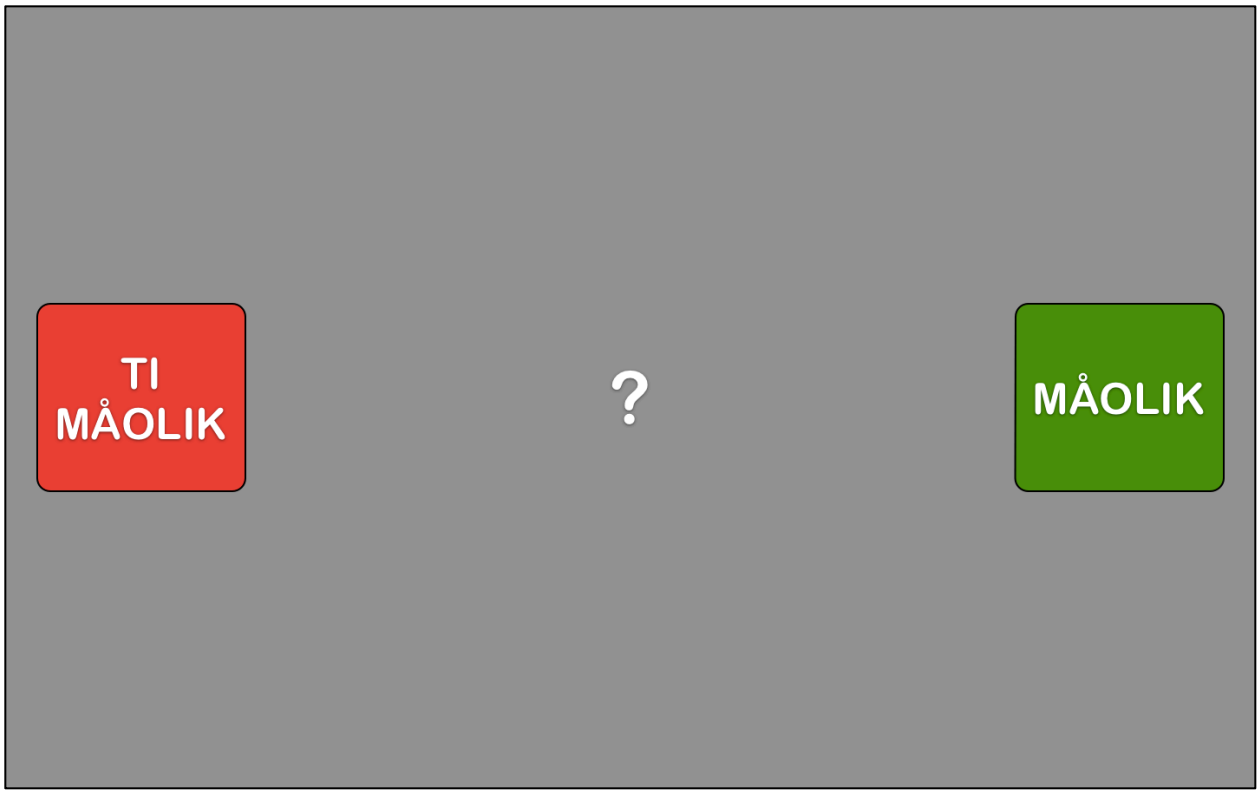

FIGURE 2. Common response screen.

The response categories *TI MÅOLIK* ('not good') and *MÅOLIK* ('good') were explained in the instruction phase. Each category was color-coded to match the sticker on the corresponding response button on the laptop keyboard. For self-paced listening, this screen appeared only after participants had advanced past the last audio segment. For preferential looking, the response screen was visible throughout the stimulus audio, and a fixation symbol was displayed in place of the question mark. At the audio offset, the fixation symbol was replaced by a question mark, which served as the cue that participants should respond.

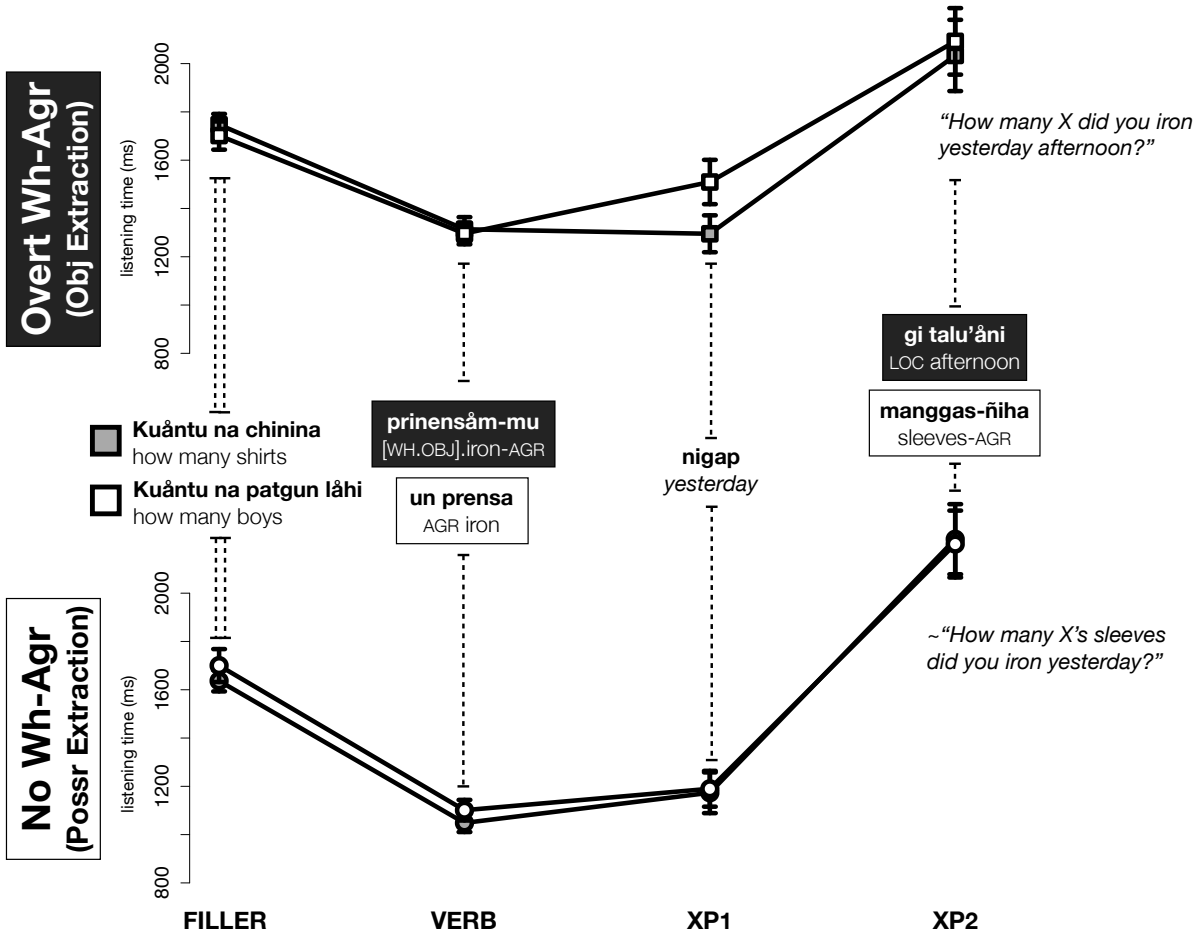

FIGURE 3. Phrase-by-phrase listening times in transitive wh-dependencies. Mean listening times, in milliseconds, are plotted above for each of the four auditory segments. Filled symbols correspond to plausible arguments; open symbols, to implausible arguments. Square symbols, top panel, correspond to object extractions (overt Wh-Agreement); round symbols, bottom panel, to possessor extractions (no Wh-Agreement). Error bars indicate standard error of the mean.

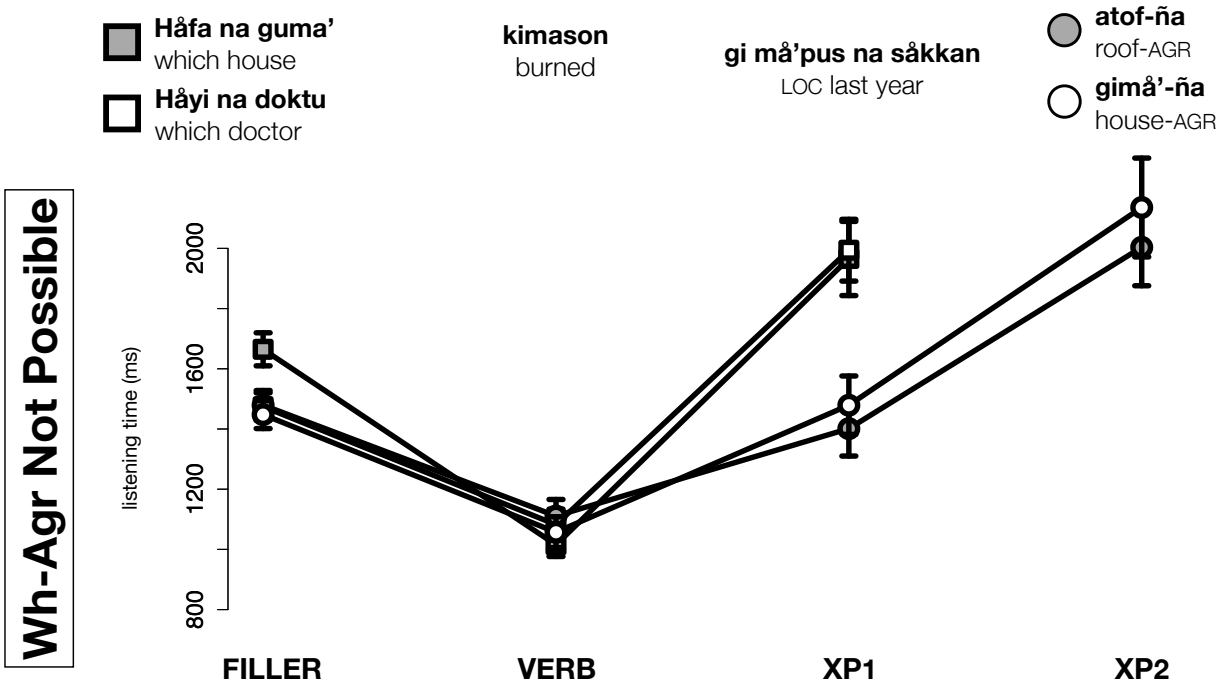

FIGURE 4. Phrase-by-phrase listening times in intransitive wh-dependencies. Mean listening times, in milliseconds, are plotted above for each of the four auditory segments. Filled symbols correspond to plausible arguments; open symbols to implausible arguments. Square symbols correspond to subject extractions; round symbols to possessor extractions. Error bars indicate standard error of the mean.

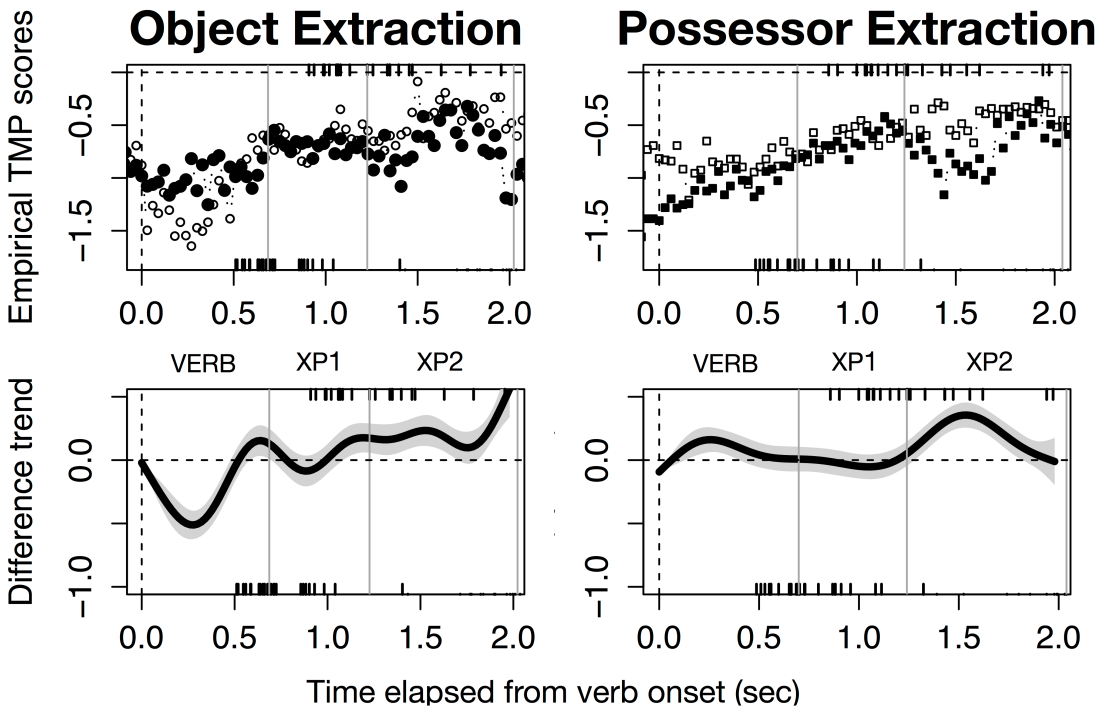

FIGURE 5. *Ti Måolik* Preference score trends for transitive wh-dependencies.

Plots to the left correspond to object extraction (overt Wh-Agreement), and plots to the right, to possessor extraction (no Wh-Agreement). Hatchmarks in plot margins indicate acoustic boundaries for the onset of *adjunct XP1* (solid, bottom margin) and the offset of *adjunct XP1* (solid, top margin). Light grey lines indicate the median of each cluster of boundaries. Median sentence offset was 2.0 sec.

Upper panel TMP scores for Plausible and Implausible Arguments - closed and open symbols, respectively.

Lower panel These plots visualize the difference between the Plausible and Implausible Argument preference scores in the corresponding upper panel plot. The thick line is the smooth Elapsed Time × Plausibility interaction term from a generalized additive model and shading indicates a Bayesian credible interval. Positive values correspond to proportionally more looks to the *Ti Maolik* category in the Implausible conditions.

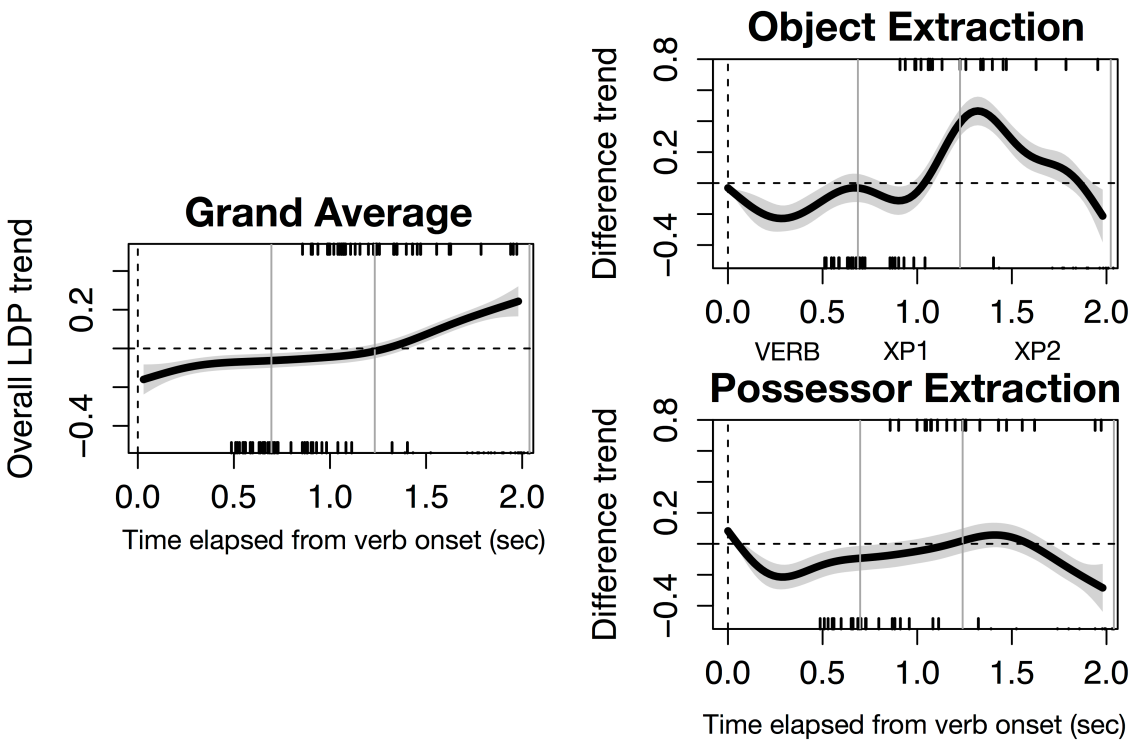

FIGURE 6. Looks-Down Preference score trends for transitive wh-dependencies.

Left panel depicts the average trend in LDP scores across all conditions. Positive values correspond to absolutely more looks down.

Right panels depict the difference in LDP scores for the Implausible versus Plausible Argument condition for each of the Extraction Types. Positive values correspond to relatively more Looks-Down for Implausible Argument conditions. The thick lines correspond to the smooth term from a generalized additive model and shading indicates a Bayesian credible interval. Tick-marks along the margin have the same interpretation as in Figure 5.

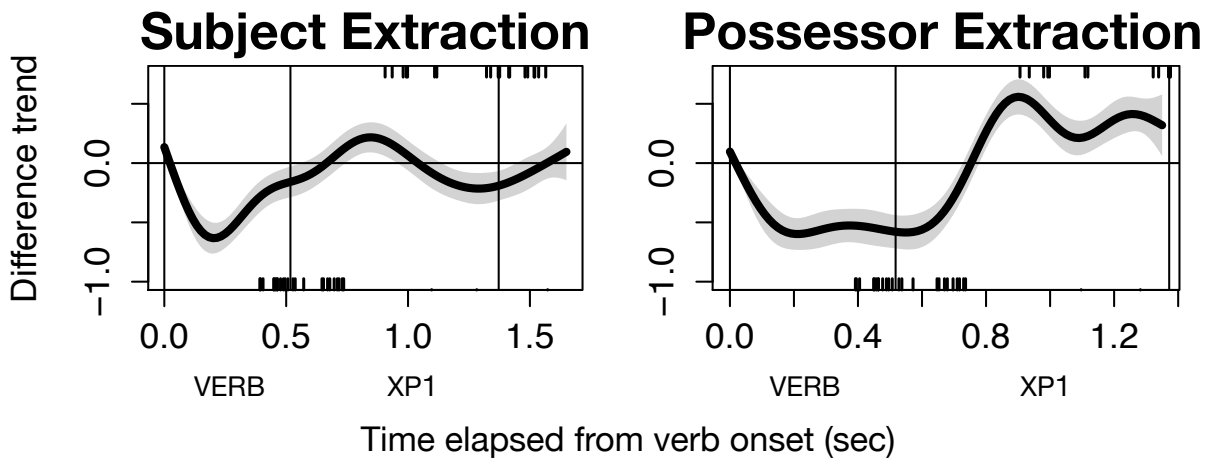

FIGURE 7. *Ti Måolik* Preference score trends for intransitive wh-dependencies.

Left panel For subject extractions, the difference in TMP scores between the Plausible and Implausible Arguments is depicted by the smooth Elapsed Time × Argument Plausibility interaction term from a generalized additive model; shading indicates a Bayesian credible interval. Positive values correspond to proportionally more looks to the *Ti Maolik* category in the Implausible conditions.

Right panel corresponds to possessor extractions. Tick-marks along the margin have the same interpretation as in Figure 5.

	Transitive Verb		**Intransitive Verb**	
	AGREEMENT	*example*	AGREEMENT	*example*
No Gap	AGR.CLITIC VERB	**un låksi**	AGR-VERB	**ekgu'**
Subject Gap	-um- + VERB	lumåksi	NO OVERT WH-AGREEMENT	**ekgu'**
Object Gap	-in- + NOMINALIZED VERB	linaksem-mu		
	NO OVERT WH-AGREEMENT	**un låksi**		
Oblique Gap	NOMINALIZED VERB	laksem-mu	NOMINALIZED VERB	ekgo'-mu
Possessor Gap	NO OVERT WH-AGREEMENT	**un låksi**	NO OVERT WH-AGREEMENT	**ekgu'**

TABLE 1. Wh-Agreement paradigm by verb type and grammatical relation of the gap.

The first row shows normal subject-verb agreement; other rows show Wh-Agreement. In each row, the agreement template is given on the left, and an inflected example on the right: *låksi* 'sew' or *ekgu'* 'jealous'. Identical realizations of Wh-Agreement are indicated in bold face. Crucially, for transitive verbs, the form used when the gap is a possessor is the same as one of the forms used when the gap is an object.

Gap is …	*Filler is …* **PLAUSIBLE**	**IMPLAUSIBLE**	*Total*
Subject (intrans)	2	2	*4*
Subject (trans)	**3**	**3**	***6***
Object	**3**	**3**	***6***
Oblique	**2**	**2**	***4***
Possessor of subject	6	2	*8*
Possessor of object	6	2	*8*
Adjunct	2	2	*4*
Total	24	16	*40*

TABLE 2. Distribution of types of gaps and fillers.

Number of items per each type of gap is cross-classified by global plausibility of the filler. Bold-face indicates conditions where the verb shows overt Wh-Agreement.

	RESIDUAL TRIMMED DATASET				UNTRIMMED DATASET			
FILLER	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	7.4	.05	160	< .001	7.4	.05	146	< .001
EXTRACTION TYPE	.05	.02	2.9	< .005	.04	.02	1.9	.07
ARG.PLAUS	-.02	.02	-1.2	.23	-.02	.02	-.97	.33
EXTR.TYPE × ARG.PLAUS	-.05	.03	-1.5	.12	-.03	.05	-.54	.59
VERB	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	7.0	.05	146	< .001	7.0	.05	134	< .001
EXTRACTION TYPE	.19	.02	10	< .001	.20	.03	6.9	< .001
ARG.PLAUS	.01	.02	.55	.58	.03	.03	.97	.34
EXTR.TYPE × ARG.PLAUS	-.01	.04	-.39	.70	-.06	.06	-1.1	.29
XP.1	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	7.0	.10	73	< .001	7.0	.10	74	< .001
EXTRACTION TYPE	.15	.03	4.9	< .001	.18	.04	4.6	< .001
ARG.PLAUS	.10	.03	3.3	< .005	.10	.04	2.5	.01
EXTR.TYPE × ARG.PLAUS	.12	.06	1.9	.05	.13	.08	1.7	.08
XP.2	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	7.5	.07	113	< .001	7.5	.07	113	< .001
EXTRACTION TYPE	-.01	.07	-.16	.87	.01	.07	.09	.92
ARG.PLAUS	.01	.07	.12	.91	-.001	.07	-.01	.99
EXTR.TYPE × ARG.PLAUS	.05	.14	.38	.70	.06	.13	.45	.65

TABLE 3. Mixed-effect models of segment listening times for transitive wh-dependencies.

Models are presented for each of the four listening segments. The 'residual-trimmed' models in the left column were derived from the 'untrimmed' models in the right column by dropping observations that stressed the untrimmed models (see section 2.4). Coefficients for Extraction Type and Argument Plausibility were centered and unit scaled to ±0.5, with Object Extractions and Implausible Fillers mapped to the positive coefficients. P-values estimated by Monte Carlo Markov Chain sampling are reported rounded to 2 digits if they are greater than .01.

	RESIDUAL TRIMMED DATASET				UNTRIMMED DATASET			
FILLER	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	7.3	.05	151	< .001	7.3	.05	152	< .001
EXTRACTION TYPE	.05	.02	3.0	< .005	.09	.03	3.2	< .005
ARG.PLAUS	-.07	.02	-4.4	< .001	-.08	.03	-2.7	<.01
EXTR.TYPE × ARG.PLAUS	-.05	.03	-1.5	.13	-.10	.06	-1.8	.07
VERB	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	6.91	.05	148	< .001	6.89	.05	144	< .001
EXTRACTION TYPE	-.01	.02	-.39	.91	-.01	.03	-.37	.71
ARG.PLAUS	-.002	.02	-.11	.70	.001	.03	.06	.95
EXTR.TYPE × ARG.PLAUS	-.06	.04	1.4	.15	.07	.06	1.1	.27
XP.1	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	7.3	.08	93	< .001	7.3	.08	92	< .001
EXTRACTION TYPE	.33	.04	8.8	< .001	.33	.05	7.0	< .001
ARG.PLAUS	.08	.04	2.1	.04	.07	.05	1.5	.13
EXTR.TYPE × ARG.PLAUS	.08	.07	1.1	.29	.02	.09	.21	.83
XP.2 (EXTR:*poss* only)	ESTIMATE	S.E.	*t*	p_{MCMC}	EST.	S.E.	*t*	p_{MCMC}
INTERCEPT	7.5	.06	116	< .001	7.4	.08	99	< .001
ARG.PLAUS	.02	.07	.34	.74	.04	.09	.48	.64

TABLE 4. Mixed-effect models of segment listening times for intransitive wh-dependencies.

Models are presented for each of the four listening segments. The 'residual-trimmed' models in the left column were derived from the 'untrimmed' models in the right column by dropping observations that stressed the untrimmed models (see section 2.4). Design coefficients for Extraction Type and Argument Plausibility factors were centered and unit scaled to ± 0.5, with Subject Extractions and Implausible Fillers mapped to the positive coefficients. P-values estimated by Monte Carlo Markov Chain sampling are reported rounded to 2 digits if they are greater than .01.

		ENDORSEMENT RATES (% judged *måolik*)		MEDIAN RESPONSE TIMES (ms; by response-type)			
		ARGUMENT	POSSESSSOR	ARGUMENT		POSSESSOR	
				MÅOLIK	TI MÅOLIK	MÅOLIK	TI MÅOLIK
TRANSITIVE	*PLAUS*	80%	66%	1198 (1402)	3036 (2496)	1877 (2984)	2631 (5675)
	IMPLAUS	21%	54%	2846 (2792)	1351 (1690)	2482 (2647)	2264 (3004)
INTRANSITIVE	*PLAUS*	81%	62%	1575 (1777)	2001 (1828)	1756 (2041)	2316 (4004)
	IMPLAUS	23%	67%	2537 (2555)	1620 (1973)	2017 (2334)	2282 (3300)

TABLE 5. Endorsement rates and reaction times from preferential looking (n = 61).

Left-hand columns. Endorsement rates, or percentage of sentences judged sensible (*måolik*), are presented by verb type (Transitive or Intransitive), extraction type (Argument [= Object or Subject] or Possessor), and the plausibility of the filler as an argument of the verb (Plausible or Implausible).

Right-hand columns. Median response times for the sensicality judgment and interquartile range (in parentheses) are presented by verb type, extraction type and plausibility. Values to the left of the dotted line correspond to response times for *måolik* responses; values to the right correspond to response times for *ti måolik* responses. Gray shading indicates the expected response.

		ENDORSEMENT RATES (% judged *måolik*)		MEDIAN RESPONSE TIMES (ms; by response-type)			
		ARGUMENT	POSSESSOR	ARGUMENT		POSSESSOR	
				måolik	*ti måolik*	*måolik*	*ti måolik*
TRANSITIVE	*PLAUS*	85%	84%	1304 (1562)	4033 (3551)	1987 (3380)	5434 (6411)
	IMPLAUS	16%	61%	2616 (2195)	1491 (2268)	2914 (2696)	2219 (2601)
INTRANSITIVE	*PLAUS*	78%	74%	1361 (1481)	2679 (1241)	1811 (1873)	4915 (4763)
	IMPLAUS	22%	69%	1816 (4062)	1503 (2294)	2142 (2257)	2530 (2712)

TABLE 6. Endorsement rates for speakers 55 years and older (n = 24).

Left-hand columns. Endorsement rates, or percentage of sentences judged sensible (*måolik*), are presented by Verb Type (transitive or intransitive), Extraction Type (argument or possessor), and the filler's plausibility as an argument of the verb (plausible or implausible).

Right-hand columns. Median response times for the sensicality judgment and interquartile range (in parenthesis) are presented by Verb Type, Extraction Type and Argument Plausibility. Values to the left of the dotted line correspond to response times for *måolik* responses; values to the right correspond to response times for *ti måolik* responses. Gray shading indicates the expected response.

			On-line				Off-line
			SELF-PACED LISTENING		PREFERENTIAL LOOKING		FORCED CHOICE
			segment				
VERB TYPE	EXTRACTION	WH-AGREEMENT	XP1	XP2	XP1	XP2	
Transitive	Object	Overt	✔	✘	✔ ✔	✔ ✔	✔
	Possessor	Possible but absent	✘	✘	✘ ✘	✔ ✘	✔
Intransitive	Subject		✔		✔ ?		✔
	Possessor		✔		✔ ?		✘

TABLE 7. Sensitivity to goodness-of-fit between filler and verb by verb type and measure

Summary of sensitivity to the goodness-of-fit between the verb and the filler, as reflected by diverse measures. On-line: listening times (for self-paced listening), and two kinds of looking preference scores (for preferential looking). Off-line: endorsement rates in forced choice sensicality judgment. Note that on-line measures at segment XP1 gauge sensitivity *before* evidence of the gap site in the input. For each measure, ✔ indicates sensitivity, ✘ indicates lack of sensitivity, and ? indicates an inconclusive result. In looking measures, the first mark corresponds to the *Ti Måolik* preference and the second, to *Looks Down* preference.