

UNIVERSITY OF ROCHESTER WORKING PAPERS IN THE LANGUAGE SCIENCES

*Center for Language Sciences, University of Rochester*

**TABLE OF CONTENTS**

PROOF-THEORETIC THEMATIC UNIQUENESS  
*Ash Asudeh* ..... 1

WHAT CHANGES WHEN WE TUNE INTO TALKER-SPECIFIC PROSODY?  
*Andrés Buxó-Lugo and Chigusa Kurumada* ..... 13

A UNIFIED OPERATOR MOVEMENT ANALYSIS FOR ADVERBIAL CLAUSES AND RELATIVE CLAUSES IN SWAHILI  
*Zhendong Liu* ..... 31

THE DIPHTHONGIZATION OBSERVATION: AN ANALYSIS OF THE “INDIE VOICE” PERSONA  
*Sarah Smith* ..... 55

THE LONG-DISTANCE *Taziji* REVISITED: EXPERIMENTAL EVIDENCE FOR INHERENT NON-LOCAL BINDINGS  
*Yuhang Xu and Jeffrey T. Runner* ..... 65



# PROOF-THEORETIC THEMATIC UNIQUENESS\*

ASH ASUDEH

*University of Rochester*

## 1 Introduction

Despite the field's long-standing interest in raising (1), a phenomenon whose study has flourished since Rosenbaum (1967), copy raising (2) has not received as much attention in theoretical linguistics.

- (1) Thora seemed to enjoy the game.
- (2) Thora seemed like she enjoyed the game.

The 'copy' pronoun in (2) is obligatory (for most speakers; Asudeh 2012: §12), which raises the question of the status of the copy raising subject's role as an argument in the semantics.

This is further complicated by the fact that copy raising, like subject-to-subject raising, alternates with an expletive-subject variant:

- (3) It seemed like Thora enjoyed the game.

In Asudeh and Toivonen (2012), we motivate a semantic role that we call PSOURCE for the copy-raised subject. We argue against using the existing thematic role STIMULUS, essentially based on the expletive alternation in (2–3), as well as broader theoretical considerations.

Our treatment of PSOURCE was inspired by a paper by Greg Carlson (Carlson 1984), a pioneering work on event semantics. In this paper, I return to Greg's paper and a problem that it discusses, Thematic Uniqueness, in light of some puzzling facts about Swedish copy raising. I first present some background on copy raising (§2.1) and on Thematic Uniqueness (§2.2). I then sketch the problem: model theory does not yield a sufficiently restrictive notion of Thematic Uniqueness (§3). I present a solution to the problem in terms of proof theory instead (§4). I conclude with some final thoughts in §5.

---

\*This paper is dedicated to Greg Carlson. It was presented at his retirement event, GregFest on May 22, 2018. Many thanks to the audience at GregFest for helpful comments and questions. And many thanks to Ida Toivonen for allowing me to write up this previously unpublished excerpt from our joint work (Asudeh and Toivonen 2012). I'm also grateful to Peter Guekguezian for his patient editorial work. Any remaining errors are my own.

## 2 Background

### 2.1 Copy Raising

Copy raising is a phenomenon in which a raising verb takes a non-expletive subject and a complement containing an obligatory pronominal ‘copy’ of the subject, as shown again here for English:

- (4) a. Thora seems like she’s found the chocolate.  
 b. \*Thora seems like Alfred’s found the chocolate.

Swedish displays a similar alternation:

- (5) a. Thora verkar som om hon har hittat chokladen.  
 Thora seems as if she has found chocolate.the  
 ‘Thora seems like she has found the chocolate.’  
 b. \* Thora verkar som om Alfred har hittat chokladen.  
 Thora seems as if Alfred has found chocolate.the

But Swedish also has the capacity to express the PSOURCE in a *på*-PP. In that case it cannot also be expressed as a subject:

- (6) a. Det verkar på Thora som om hon har hittat chokladen.  
 It seems on Thora as if she has found chocolate.the  
 ‘Thora seems like she has found the chocolate.’  
 b. \* Thora verkar på Isak/Thora som om hon har hittat chokladen.  
 Thora seems on Isak/Thora as if she has found chocolate.the

English copy raising was first discussed extensively in work by Rogers (1971, 1972, 1973, 1974), although it did receive a brief mention on the first major extended work on raising Postal (1974: 268, fn.1). To our knowledge, Asudeh and Toivonen (2012) is the first work to discuss Swedish copy raising in any detail. For further references on both English and Swedish copy raising, see Asudeh and Toivonen (2012, 2017) and Toivonen (2020).

### 2.2 Thematic Uniqueness

Carlson (1984: 270–273) discusses the status of thematic roles in the grammar and issues raised by the apparent universal constraint against verbs like the made up verb *skick*, which takes a subject that is an AGENT and two objects that are both LOCATION (or PATIENT — the exact thematic role doesn’t matter).

- (7) John skicked Bill’s leg Bill’s shin. (Carlson 1984: 271, (11))  
 (Meaning: John kicked Bill on the shin part of his leg)

Carlson (1984) concludes that there must be some principle of Thematic Uniqueness, which states that any eventuality can have at most one instance of any particular thematic role. Carlson further argues that Thematic Uniqueness should not be captured either strictly semantically or syntactically, but is rather a constraint on eventualities (Carlson 1984: 273).

One way of capturing Thematic Uniqueness is through a principle such as the following:.

- (8)        **Unique Role Requirement** (Landman 2000: 38)  
 If a thematic role is specified for an event, it is uniquely specified.

The question is how to operationalize this.

There is a long tradition in the literature that captures the Unique Role Requirement in the model theory by defining thematic roles as partial functions from eventualities to individuals (Chierchia 1984, 1989, Landman 2000, Champollion 2015, 2017). The underlying explanation for the ill-formedness of (9) — which is an attempt to use *kick* as if it were the hypothetical *skick* — and (10) is thus potentially the same:

- (9)        \* John kicked Bill's leg Bill's shin.  
 (10)       \* Tom verkar på Robin som om han skrattar.  
              Tom seems on Robin as if he laughs

Carlson assumes that (9) involves an attempt to assign two instances of a LOCATION thematic role and therefore violates the requirement that each event have at most one instance of a given role. Under the model-theoretic treatment of the uniqueness requirement, this follows from the definition of thematic roles as functions. Similarly, (10) involves two instances of the PSOURCE function and is blocked for the same reason.

### 3 The Problem: Model Theory is Too Weak

There is, however, reason to believe that denotational or model-theoretic uniqueness is not a sufficiently strong uniqueness requirement. In particular, model-theoretic uniqueness makes false empirical predictions in cases of denotational equivalence, a fact that was already foreseen by Carlson (1984: 272). Foreshadowing, the solution is to therefore introduce a proof-theoretic notion of uniqueness that guarantees uniqueness irrespective of denotation. It is important to realize that the two notions do not conflict: Proof-theoretic uniqueness is independent of the model-theory and it can therefore supplement model-theoretic denotation rather than necessarily supplanting it. Furthermore, proof-theoretic uniqueness does not introduce any new mechanisms into the theory, since it relies only on the mechanism that already regulates functor-argument composition and, specifically, proper predicate saturation and argument consumption.

Consider example (10) if *Tom* and *Robin* are different names for the same individual.

- (10)       \* Tom verkar på Robin som om han skrattar.  
              Tom seems on Robin as if he laughs

If the names are denotationally equivalent, then there is no violation of model-theoretic uniqueness, since the PSOURCE function is simply returning two different instances of the same individual. This particular case is perhaps not too worrying, though, since it could potentially be handled through the use of intensions (depending on how the semantics of proper names is treated).

A potentially more problematic case is where there are two occurrences of the same individual using the same name, which is equally ill-formed:

- (11)       \* Tom verkar på Tom som om han skrattar.  
              Tom seems on Tom as if he laughs

Here there is very strong denotational equivalence between the two individuals designated by *Tom* and model-theoretic uniqueness breaks down. It may be, however, that independent factors account for the ill-formedness of (11). First, there is a potential Principle C violation in terms of binding theory (Chomsky 1981), although given the weakness of Principle C (Evans 1980, 1985) this seems like a tenuous explanation for the strong ungrammaticality of (11). One might instead contend that repeated uses of the same name use different guises (Castañeda 1972, 1989, Heim 1998). This might also explain the ungrammaticality of (10).

However, none of these explanations can readily account for the ungrammaticality of a parallel example with a reflexive in the *på*-PP:

- (12) \* Tom verkar på sig själv som om han skrattar.  
Tom seems on himself as if he laughs

The reflexive in this example is not logophoric or special in any way: It must be denotationally equivalent to its antecedent. Yet the sentence is ill-formed despite the model-theoretic/denotational equivalence of the subject and the adjunct.

We are therefore led to reconsider denotational uniqueness in the model as the means of capturing the Unique Role Requirement, which is repeated here:

- (13) **Unique Role Requirement** (Landman 2000: 38)  
If a thematic role is specified for an event, it is uniquely specified.

The conclusion reached here is that, in generalizing this notion to PSOURCES, the expression “uniquely specified” in (13) cannot be understood as “having a unique denotation in the model”, which is what the usual functional understanding of thematic roles captures.

The data in (10–12) indicates that — when extended to eventuality participants in general, including PSOURCES — the uniqueness requirement is that eventuality participants are uniquely *overtly* realized, i.e. that they have at most one syntactic realization. Yet Carlson (1984: 270–272) cautions us against understanding thematic roles as purely syntactic constructs, as they seem to do no purely syntactic work. And, in any case, the Swedish *på*-PP does not seem to be a syntactic argument (Asudeh and Toivonen 2012).

In sum, it is insufficient to capture the uniqueness requirement solely in terms of the model-theoretic semantics. Rather, it must somehow be stated as a condition on the mapping from the syntax to the semantics.

## 4 A Solution: Proof-Theoretic Uniqueness

A method for correctly capturing the uniqueness requirement becomes apparent if we consider more basic cases of thematic roles with respect to uniqueness:<sup>1</sup>

- (14) \* Tom skrattar sig själv.  
Tom laughs him self

<sup>1</sup>I use a Swedish example to ensure that the reflexive is not understood emphatically, as it could be in the equivalent English sentence (on a par with *Tom himself laughs*). The emphatic reflexive in Swedish would just be *själv*, without *sig*.

The model-theoretic treatment of the uniqueness requirement on thematic roles does not block this sentence, since the subject and the reflexive are denotationally equivalent. They could both be assigned the thematic role AGENT, for example. In theory-neutral terms, what actually blocks (14) is instead whatever ensures proper predicate saturation or, equivalently, argument consumption. The sentence is bad because *skratta* ('laugh') takes only one argument and the second argument cannot be handled properly.

It is instructive to consider this in light of the Theta Criterion (Chomsky 1981: 36):

(15) **Theta Criterion**

Each argument bears one and only one  $\theta$ -role and each  $\theta$ -role is assigned to one and only one argument.

Chierchia (1984) explicitly relates the uniqueness requirement on thematic roles to the Theta Criterion. He argues that treating thematic roles as functions captures the second part of the Theta Criterion. We have seen that this is in fact only true up to denotational equivalence. However, it is if anything the first part of the Theta Criterion that blocks (14): The object does not bear a theta role, since the sole theta role of the verb is assigned to the subject.

But in subsequent work in the Minimalist Program, the first clause of the Theta Criterion was abandoned (Brody 1993, Bošković 1994, Hornstein 1999) and Chomsky (1995: 200) has argued that the entire principle can be subsumed under Full Interpretation (FI). If FI is to be construed as a restriction on semantic interpretation or the mapping from Logical Form to the Conceptual–Intentional System, it cannot explain the ungrammaticality of (12), repeated here, since it receives a perfectly valid (if redundant) compositional interpretation:

(16) \* Tom verkar på sig själv som om han skrattar.  
Tom seems on himself as if he laughs

(17)  $\exists s.\text{seem}(s, \exists e[\text{laugh}(e, \text{tom}) \wedge \text{AGENT}(e) = \text{tom}]) \wedge \text{PSOURCE}(s) = \text{tom} \wedge \text{PSOURCE}(s) = \text{tom}$

Here  $s$  is a variable over states and  $e$  is a variable over events; see Asudeh and Toivonen (2012: 371–372) for details.

Asudeh has previously argued that Full Interpretation can be reduced to a proof-theoretic notion of *resource sensitivity* (Asudeh 2004: 99ff., Asudeh 2012: §5). Resource sensitivity is captured through the use of a resource logic for semantic composition, as in Glue Semantics (Dalrymple 1999, 2001, Dalrymple et al. 2019, Asudeh 2012), which uses the resource logic *linear logic* (Girard 1987) for composition. Premises in linear logic proofs are resources whose use is tightly controlled: A successful linear logic proof requires each premise to be used exactly once.

Asudeh (2012) calls the resource sensitivity that stems purely from properties of the underlying logic *Logical Resource Sensitivity*.

(18) **Logical Resource Sensitivity**

In a resource logic, premises in proofs cannot be freely *reused* or *discarded*.

A related notion that is more useful for linguistics, *Linguistic Resource Sensitivity*, is derived by stating a linguistically motivated goal condition on the linear logic proof for semantic composition (Asudeh 2012: 106–110).

(19) **Linguistic Resource Sensitivity**

Elements of combination in grammars cannot be freely *reused* or *discarded*.

In the absence of such a goal condition, the premises could be properly used up by simply conjoining them all together, but this does not derive a properly composed meaning.

A typical goal condition in Glue Semantics is the following:

$$(20) \quad \Gamma \vdash \phi : s$$

From a premise set  $\Gamma$ , the goal is to establish an atomic conclusion  $s$  that corresponds to the interpretation of the sentence, represented as  $\phi$ .<sup>2</sup>

On this view, (14), repeated here, is ill-formed because there are resources contributed by the subject and object, but the verb only consumes the subject resource, illicitly leaving behind the object resource.

$$(21) \quad * \text{ Tom skrattar sig själv.} \\ \text{Tom laughs him self}$$

This is schematized in the following proof ( $\multimap$  is linear implication and  $\otimes$  is the relevant kind of linear conjunction):

$$(22) \quad \frac{\frac{\text{SUBJECT} \quad \text{SUBJECT} \multimap \text{VERB}}{\text{VERB}} \quad \text{OBJECT}}{\text{VERB} \otimes \text{OBJECT}}$$

Notice that argument consumption corresponds to implication elimination. The goal condition (20) is not met, since the result is a conjunction, not an atomic term.

This suggests a way to address the problem of denotational uniqueness by replacing the model-theoretic version of the uniqueness requirement with a proof-theoretic version. The basic idea is to extend the calculus of argument consumption to PSOURCES, but without necessarily treating them as arguments (which we argue against in Asudeh and Toivonen 2012). This is accomplished by embedding the meanings for raising verbs and the *på*-PP adjuncts in a Glue Semantics analysis that introduces a PSOURCE resource in the linear logic term for semantic composition. Linguistic Resource Sensitivity will then yield a proof-theoretic uniqueness requirement that works regardless of denotation. Note that this proof-theoretic treatment does not conflict with model-theoretic uniqueness and I will continue to assume that PSOURCES and thematic roles are partial functions.

Rather than attempting to demonstrate proof-theoretic uniqueness in the abstract, let us work through the relevant cases. First let us consider subject-to-subject raising and expletive examples in English and Swedish:

$$(23) \quad \begin{array}{ll} \text{a.} & \text{Tom seems to be laughing.} \\ \text{b.} & \text{Tom verkar skratta.} \\ & \text{Tom seems laugh.INF} \\ & \text{‘Tom seems to be laughing.’} \end{array}$$

<sup>2</sup>The linear logic terms are also typed and the type of the conclusion here is  $t$ . I leave typing aside here, since it is not relevant to the point at hand.

- (24) a. It seems like Tom is laughing.  
 b. Det verkar som om Tom skrattar.  
 It seems as if T. laughs  
 ‘It seems as if Tom is laughing.’

The interpretations for these cases are presented here as two separate terms (whose proper interaction is captured in the Glue logic side, presented in (26) below):

$$(25) \quad \lambda p \lambda s'. \mathbf{seem}(s', p) \\ \lambda S \lambda s. \exists v [S(s) \wedge \mathbf{PSOURCE}(s) = v]$$

Asudeh and Toivonen (2012) argue that the existential closure is obligatory in both English and Swedish subject-to-subject raising and in English expletive examples. The closure is only optional in Swedish expletive examples, to allow composition with a *på*-PP adjunct.

These interpretations are embedded in Glue meaning constructors, which pair terms of the meaning language with linear logic terms:

$$(26) \quad \lambda p \lambda s'. \mathbf{seem}(s', p) : \\ \text{COMPLEMENT} \multimap \text{PSOURCE} \multimap \text{EVENT} \multimap \text{RESULT} \\ \lambda S \lambda s. \exists v [S(s) \wedge \mathbf{PSOURCE}(s) = v] : \\ (\text{PSOURCE} \multimap \text{EVENT} \multimap \text{RESULT}) \multimap (\text{EVENT} \multimap \text{RESULT})$$

The linear logic terms are provided schematically here, but normally they would be instantiated in terms of some syntactic theory, such as Lexical-Functional Grammar (Bresnan et al. 2016).

Crucially, a linear logic term is introduced for the PSOURCE. This will serve as a resource that must be properly consumed in the linear logic proof. The other linear logic terms stand for the raising verb’s sentential complement, the event variable, and the result of composition. It is important to bear in mind that linear logic terms with identical names in proofs are meant to be understood as token-identical. The composition of the examples in (23–24) proceeds as in (27), leaving aside details of the complement (for details, see Asudeh and Toivonen 2012). The linear logic are here abbreviated terms in order to save space.

$$(27) \quad \begin{array}{c} \vdots \\ \lambda p \lambda s'. \mathbf{seem}(s', p) : \quad \mathbf{laugh}(\dots) : \\ C \multimap P \multimap E \multimap R \quad C \\ \hline \lambda s'. \mathbf{seem}(s', \mathbf{laugh}(\dots)) : \quad \lambda S \lambda s. \exists v [S(s) \wedge \mathbf{PSOURCE}(s) = v] : \\ P \multimap E \multimap R \quad (P \multimap E \multimap R) \multimap (E \multimap R) \\ \hline \lambda S. \exists s [S(s)] : \quad \lambda s. \exists v [\mathbf{seem}(s, \mathbf{laugh}(\dots)) \wedge \mathbf{PSOURCE}(s) = v] : \\ (E \multimap R) \multimap R \quad E \multimap R \\ \hline \exists s \exists v [\mathbf{seem}(s, \mathbf{laugh}(\dots)) \wedge \mathbf{PSOURCE}(s) = v] : R \end{array}$$

Notice that implication elimination in the linear logic corresponds to functional application in the meaning language (via the Curry-Howard isomorphism; Curry and Feys 1958, Howard 1980). I assume standard existential closure of the matrix event variable  $s$  in the absence of other quantification.

The crucial step is the one where the existential closure  $\exists v$  applies to the raising verb's meaning. On the linear logic side, the existential closure needs to consume an implication from a PSOURCE, which is provided by the raising verb. This is a standard higher-type functor-argument application for a quantifier combining with its scope. The final result is an atomic linear logic term corresponding to the sentential semantics, which satisfies the goal condition (20) above.

Let us next see the meaning constructor for the Swedish preposition *på* in a *på*-PP PSOURCE adjunct:

$$(28) \quad \lambda x \lambda S \lambda s. S(s) \wedge \text{PSOURCE}(s) = x : \\ \text{OBJECT} \multimap \\ (\text{MODIFIEE'S PSOURCE} \multimap \text{MODIFIEE'S EVENT} \multimap \text{MODIFIEE'S RESULT}) \multimap \\ (\text{MODIFIEE'S EVENT} \multimap \text{MODIFIEE'S RESULT})$$

The Glue logic side treats the adjunct as a modifier on a term that depends on a PSOURCE. In this respect, the *på*-adjunct is like the existential closure term in (26). Both contribute linear logic terms that want to consume a dependency on a PSOURCE.

This is sufficient to explain the ungrammaticality of Swedish subject-to-subject raising with a *på*-PP:

$$(29) \quad * \text{ Tom verkar på Sara skratta.} \\ \text{T. seems on S. laugh.INF}$$

The existential closure of PSOURCE is obligatory in this case. This means that both the existential closure meaning constructor and the *på*-PP meaning constructor are seeking to consume a term of the form PSOURCE  $\multimap$  EVENT  $\multimap$  RESULT. However, only one instance of this term has been contributed by the verb. The resource sensitivity of linear logic entails that once one of these PSOURCE consumers has consumed the dependency on the verb's PSOURCE, there is no way to satisfy the other consumer.

This is shown schematically in the following packed proof:

$$(30) \quad \begin{array}{c} \vdots \\ \text{raising verb} \quad \text{complement} \\ C \multimap P \multimap E \multimap R \quad C \\ \hline \exists\text{-clos./p\AA-PP} \quad P \multimap E \multimap R \quad \exists\text{-clos./p\AA-PP} \\ (P \multimap E \multimap R) \multimap (E \multimap R) \quad (P \multimap E \multimap R) \multimap (E \multimap R) \\ \hline (P \multimap E \multimap R) \multimap (E \multimap R) \quad E \multimap R \\ \hline [(P \multimap E \multimap R) \multimap (E \multimap R)] \otimes (E \multimap R) \end{array}$$

It is readily apparent that the final result is not an atomic term and that (20) is therefore not satisfied.

Lastly, let us consider copy raising:

$$(31) \quad \begin{array}{ll} \text{a.} & \text{Tom seems like he is laughing.} \\ \text{b.} & \text{Tom verkar som om han skrattar.} \\ & \text{T. seems as if he laughs} \\ & \text{'Tom seems as if he is laughing.'} \end{array}$$

In both English and Swedish, the copy-raised subject serves as the PSOURCE.

This is captured by embedding the copy raising verb's meaning in the following meaning constructor:

$$(32) \quad \lambda x \lambda P \lambda s. \mathbf{seem}(s, P(x)) \wedge \text{PSOURCE}(s) = x : \\ \text{SUBJECT/PSOURCE} \multimap (\text{SUBJECT} \multimap \text{COMPLEMENT}) \multimap \text{EVENT} \multimap \text{RESULT}$$

Notice that for non-expletive variants (i.e., true copy raising), there is no existential closure, since we want the matrix copy raising subject to be the PSOURCE. In the syntactic analysis of Asudeh (2012) and Asudeh and Toivonen (2012), the copy-raised subject is structure-shared with the implicit subject of the predicative *like/som* complement and the copy raising verb composes its subject with the property corresponding to its complement. From a resource-logical perspective, the important aspect of the linear logic term in (32) is that the only consumer of the matrix subject/PSOURCE is the copy raising verb. If a *på*-PP modifies a copy raising verb, there are two possible proofs, but neither terminates in an atomic linear logic term.

The first possibility is if the copy raising verb composes with its subject directly. There is then no dependency on a PSOURCE left in the proof and the *på*-PP modifier cannot find its scope. This is shown schematically here:

$$(33) \quad \begin{array}{c} \text{subject} \quad \text{CR verb} \quad \vdots \\ S \quad S/P \multimap (S \multimap C) \multimap E \multimap R \quad \text{complement} \\ \hline (S \multimap C) \multimap E \multimap R \quad \text{property} \\ S \multimap C \quad \text{på-PP} \\ \hline E \multimap R \quad (P \multimap E \multimap R) \multimap (E \multimap R) \\ \hline (E \multimap R) \otimes (P \multimap E \multimap R) \multimap (E \multimap R) \end{array}$$

Alternatively, if the *på*-PP adjunct consumes the dependency on the PSOURCE, then there is no longer a consumer for the matrix subject, since the subject and the PSOURCE are one and the same. This is shown schematically in the following proof, where the term for the copy raising verb has been curried to compose with the complement first, for ease of presentation:

$$(34) \quad \begin{array}{c} \vdots \\ \text{complement} \\ \text{property} \\ S \multimap C \\ \hline \text{CR verb} \\ (S \multimap C) \multimap S/P \multimap E \multimap R \\ \hline P \multimap E \multimap R \\ \text{på-PP} \\ (P \multimap E \multimap R) \multimap (E \multimap R) \\ \hline \text{subject} \\ S \quad E \multimap R \\ \hline S \otimes (E \multimap R) \end{array}$$

A *på*-PP adjunct therefore cannot co-occur with a copy raising verb for proof-theoretic reasons: There are not enough instances of the subject/PSOURCE to satisfy all consumers (the copy raising verb and the adjunct).

This proof-theoretic treatment of PSOURCE uniqueness is entirely independent of denotations and depends solely on the linear logic terms for semantic composition. Proof-theoretic uniqueness therefore blocks all instances of copy raising with *på*-PP adjuncts, including the denotationally equivalent instances in (10–12) above and particularly the pernicious reflexive case, repeated here:

- (35) \* Tom verkar på sig själv som om han skrattar.  
 T. seems on him self as if he laughs

This example is ill-formed for the proof-theoretic reasons just outlined with respect to the proofs (33) and (34) above.

Since Glue proofs are essentially structural representations of the syntax-semantics interface (Asudeh and Crouch 2002a,b), proof-theoretic uniqueness therefore has the desired property of controlling for the linguistic realization of PSOURCES through the mapping from syntax to semantics, based on the resources underlying contributions of PSOURCE, rather than controlling for denotational equivalence in the model theory.

## 5 Conclusion

The basis for proof-theoretic uniqueness is Linguistic Resource Sensitivity, which controls proper argument consumption by predicates. I argued that it is this latter notion that could be responsible for blocking cases involving thematic roles that denotational uniqueness lets slip through, such as unlicensed reflexives. The proof-theoretic control of functor-argument combination effected by Linguistic Resource Sensitivity was generalized to PSOURCES by assigning them a resource that must be properly consumed in the proof, although in the model-theoretic semantics they are still not treated as arguments.

Proof-theoretic uniqueness is thus a stronger condition than model-theoretic uniqueness, although the independence of the two kinds of uniqueness means that there is no conflict between the two and they can be captured simultaneously in one system, as they have been here, since I still assume that thematic/semantic roles are functions on eventualities, as per Carlson's original pioneering insight.

One problem remains, however. Normally, a strong correspondence is assumed between proofs and models, as captured by the Curry-Howard Isomorphism (Curry and Feys 1958, Howard 1980). The solution sketched here puts some stress on that correspondence if it is construed as a correspondence between the terms in the meaning language and the terms in the Glue logic. It is worth pointing out, though, that the correspondence is preserved between the compositional structure of the proof itself (as captured in proof rules) and the models for the proofs.

## References

- Asudeh, Ash. 2004. Resumption as resource management. Doctoral Dissertation, Stanford University.
- Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford: Oxford University Press.
- Asudeh, Ash, and Richard Crouch. 2002a. Coordination and parallelism in Glue Semantics: Integrating discourse cohesion and the element constraint. In *Proceedings of the LFG02 conference*, ed. Miriam Butt and Tracy Holloway King, 19–39. Stanford, CA: CSLI Publications.
- Asudeh, Ash, and Richard Crouch. 2002b. Derivational parallelism and ellipsis parallelism. In *Proceedings of the 21st West Coast Conference on Formal Linguistics*, ed. Line Mikkelsen and Christopher Potts, 1–14. Somerville, MA: Cascadilla Press.

- Asudeh, Ash, and Ida Toivonen. 2012. Copy raising and perception. *Natural Language and Linguistic Theory* 30:321–380.
- Asudeh, Ash, and Ida Toivonen. 2017. A modular approach to evidentiality. In *Proceedings of the LFG14 conference*, ed. Miriam Butt and Tracy Holloway King, 45–65. Stanford, CA: CSLI Publications.
- Bošković, Željko. 1994. D-structure, theta-criterion, and movement into theta-positions. *Linguistic Analysis* 24:247–286.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2016. *Lexical-functional syntax*. Malden, MA: Wiley-Blackwell, 2 edition.
- Brody, Michael. 1993.  $\theta$ -theory and arguments. *Linguistic Inquiry* 24:1–23.
- Carlson, Gregory N. 1984. Thematic roles and their role in semantic interpretation. *Linguistics* 22:259–279.
- Castañeda, Héctor-Neri. 1972. Thinking and the structure of the world: *Discours d'ontologie. Crítica: Revista Hispanoamericana de Filosofía* 6:43–86.
- Castañeda, Héctor-Neri. 1989. *Thinking, language, and experience*. Minneapolis: University of Minnesota Press.
- Champollion, Lucas. 2015. The interaction of compositional semantics and event semantics. *Linguistics and Philosophy* 38:31–66.
- Champollion, Lucas. 2017. *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Oxford: Oxford University Press.
- Chierchia, Gennaro. 1984. Topics in the syntax and semantics of infinitives and gerunds. Doctoral Dissertation, University of Massachusetts, Amherst.
- Chierchia, Gennaro. 1989. Structured meanings, thematic roles and control. In *Properties, types and meaning. volume II: Semantic issues*, ed. Gennaro Chierchia, Barbara H. Partee, and Raymond Turner, 131–166. Dordrecht: Kluwer.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Curry, Haskell B., and Robert Feys. 1958. *Combinatory logic*, Vol. 1. Amsterdam: North-Holland.
- Dalrymple, Mary, ed. 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. Cambridge, MA: MIT Press.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.
- Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The oxford reference guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

- Evans, Gareth. 1980. Pronouns. *Linguistic Inquiry* 11:337–362. Reprinted in Evans (1985: 214–248).
- Evans, Gareth. 1985. *Collected papers*. Oxford: Clarendon Press. Edited by Antonia Phillips.
- Girard, Jean-Yves. 1987. Linear logic. *Theoretical Computer Science* 50:1–102.
- Heim, Irene. 1998. Anaphora and semantic interpretation: A reinterpretation of Reinhart’s approach. In *The interpretive tract*, ed. Uli Sauerland and Orin Percus, 205–246. Cambridge, MA: MITWPL.
- Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30:69–96.
- Howard, William A. 1980. The formulae-as-types notion of construction. In *To H.B. Curry: Essays on combinatory logic, lambda calculus and formalism*, ed. Jonathan P. Seldin and J. Roger Hindley, 479–490. London: Academic press. Circulated in unpublished form from 1969.
- Landman, Fred. 2000. *Events and plurality: the Jerusalem lectures*. Dordrecht: Kluwer.
- Postal, Paul. 1974. *On raising*. Cambridge, MA: MIT Press.
- Rogers, Andy. 1971. Three kinds of physical perception verbs. In *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*, 206–222.
- Rogers, Andy. 1972. Another look at flip perception verbs. In *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*, 303–315.
- Rogers, Andy. 1973. Physical perception verbs in English: A study in lexical relatedness. Doctoral Dissertation, UCLA.
- Rogers, Andy. 1974. A transderivational constraint on Richard? In *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*, 551–558.
- Rosenbaum, Peter S. 1967. *The grammar of English predicate complement constructions*. Cambridge, MA: MIT Press.
- Toivonen, Ida. 2020. Perception verbs and evidentiality in Swedish and English. Ms., Carleton University.

# WHAT CHANGES WHEN WE TUNE INTO TALKER-SPECIFIC PROSODY?

ANDRÉS BUXÓ-LUGO

CHIGUSA KURUMADA

*Department of Psychology, University at Buffalo*

*Department of Brain and Cognitive Sciences, University of Rochester*

## Abstract

A major puzzle of spoken language comprehension is how listeners navigate variability in speech prosody to infer the intended meaning, for instance pitch movements signaling questions vs. statements. Listeners must adapt to variability across talkers (e.g., according to their age, gender, accents) as well as linguistic contexts. The relationship between the phonetic variations and linguistic (e.g., syntactic) contexts that host the variations is not yet well-understood, however. Using resynthesized 11-step continua between rising and falling contours, we investigated the extent to which prosodic adaptation is conditioned on a syntactic context. Over two blocks, participants repeatedly categorized “It’s X-ing” items (e.g., “It’s cooking”) sampled from the continua as questions or statements. In between, one group of listeners heard ambiguous items midway along the continua with different, interrogative syntax (“Is it cooking”) and items near the falling end of the continua with declarative syntax (“It’s cooking”). Another group heard the reverse. Prosodic adaptation was observed only in the listeners who heard the ambiguous prosody with the declarative syntax, congruent with the test tokens. We argue that the effective contextual conditioning facilitates robust prosodic adaptation while offsetting the risk of over-generalization across tokens with distinct phonetic features.

## 1 Introduction

Human language allows us to communicate highly complex and abstract information such as intentions and emotions. Speech prosody, defined as a holistic impression of melodic and rhythmic aspects of speech, plays a critical role in this process (e.g., Bolinger, 1964; Pierrehumbert & Hirschberg, 1990; Ladd, 2008; Cole, 2015; Cutler, 2015; Dahan, 2015). Prominence placed on a particular word can convey the talker's intention to signal a contextually salient contrast (e.g., "Olivia likes LEMONS (not oranges)" vs. "OLIVIA (not Annie) likes lemons"). Similarly, a rising terminal pitch of an utterance can distinguish a question from a statement. Complicating the listeners' task, however, realizations of prosody can vary substantially across talkers due to factors both systematic (e.g., vocal tract length, speaking rate) and incidental (e.g., speech errors) (Arvaniti, 2011; D'Imperio, Grice & Cangemi, 2016; Breen et al, 2018). Understanding the meaning of prosody must therefore include processes that aid listeners in navigating this talker-variability.

Recent studies have begun to uncover sources of flexibility necessary for robust prosodic processing. Neuroimaging work has suggested that the human auditory cortex encodes talker-normalized *relative* pitch in addition to absolute pitch (Tang et al., 2017). Furthermore, listeners' interpretations of coherent pitch movements (i.e., intonation) can further adapt to the statistics of the acoustic / phonetic features of a given talker. For instance, Xie, Buxó-Lugo, & Kurumada (2021) exposed three groups of listeners to distinct statistics of pitch movements signaling questions or statements. After only eight minutes of exposure, these groups of listeners derived diverging interpretations for novel, ambiguous tokens. Similar cases of rapid prosodic adaptation have been demonstrated for pragmatic (Patel et al., 2011; Ito et al, 2017; Kurumada et al., 2017; Nakamura et al., 2019) and affective (Woodard et al., 2021) meanings. What is as of yet mostly unknown is: *What changes in response to an accumulating exposure to a particular talker?*

The received wisdom in the literature suggests that prosodic processing calibrates to, or compensates for, a base vocal pitch and its range for a given talker. Beckman (1995) puts it as "all of our theories of intonational structure include at least an implicit representation of the speaker's overall pitch range in our models of the hearer's competence." Listeners may apply distinct expectations for vocal pitch in, say, female vs. male talkers (Gussenhoven & Rietveld, 1998; Bishop & Keating, 2003) or children vs. adults (Patel & Brayton, 2009). And such adjustment can happen rapidly (Lee, 2009). In real time as an utterance unfolds, an earlier part of an utterance serves as an anchor to shift expectations for subsequent parts (e.g., Dilley & Pitt, 2010; Saindon et al., 2017). As a result, listeners can flexibly map specific stimulus values to linguistically meaningful prosodic categories (e.g., rising vs. falling pitch) at different rates across talkers.

What has not been directly tested is whether listeners adapt more specific, *phonetic* knowledge of a prosodic category beyond just an overall baseline or range of a given cue (e.g., F0). For instance, a "question" prosody is hardly a homogeneous category either in its form or its meaning (Gunlogson, 2003; Hedberg et al., 2014). It has been argued that a question with declarative syntax (e.g., "It's raining?") is typically associated with a larger degree of pitch excursion compared to one with interrogative syntax (e.g., "Is it raining?") even when they are deemed equivocal in their phonological properties (Haan, 2001). Such systematic, phonetic variations may mean that not only do listeners need to adapt to talkers' base vocal pitch and pitch range, but they may also need to tune into *how these cross-talker variations might interact with*

*syntactic contexts* and subtle meaning differences associated with them (e.g., How do Talker A's declarative (vs. interrogative) questions differ from Talker B's?).

Functionally, optimal solutions for listeners can be affected by two motivations. On the one hand, listeners need to adapt and generalize the adaptation efficiently. Characteristics of a talker seen in one type of question should generalize to another (unseen) type of question, or else talker-adaptation would take too long to be practically useful for communication. On the other hand, listeners must adapt effectively and perhaps conservatively. Pitch movements observed for a declarative question, for example, might not directly predict how the same talker would produce an interrogative question. Such a balancing act has indeed been observed for adaptation to pronunciation variations in phonemes (e.g., vowels and consonants) and their positions in a word (Ades, 1974; Samuel, 1989) or lexical identity (Dahan & Mead, 2010). For example, stimulus features of a word-initial /d/ (e.g., *date*) and a word-final /d/ (e.g., *paid*) are typically distinct from each other, and exposure to one does not automatically produce talker-specific adaptation in the other (Samuel, 2020). The knowledge of fine-grained, context-dependent variations is thus thought to help reduce the risk of wholesale adaptation that is not warranted by natural distributions in the input. The current study asked if a similar context-sensitive conditioning would occur in prosodic adaptation.

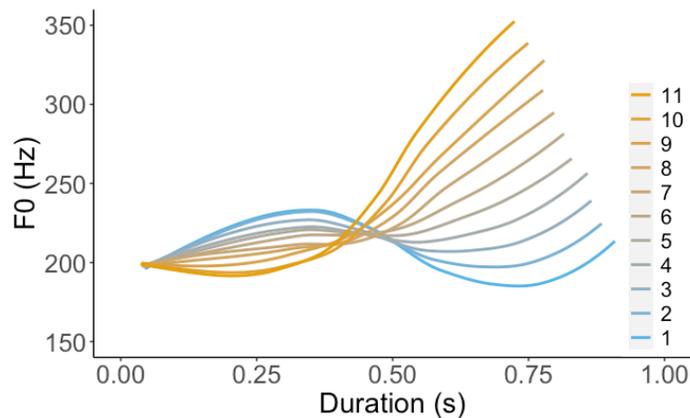
## 2 Methods

### 2.1 Participants

180 participants were recruited through the online platform Amazon Mechanical Turk (<https://www.mturk.com/>). Five participants were excluded for providing a uniform response (either question or statement) for all tokens, leaving 175 participants for the analysis. Participants were all self-identified native speakers of American English and received monetary compensation (\$4.00) for their participation.

### 2.1 Stimuli

The test stimuli were the same as those used in Xie, Buxó-Lugo, and Kurumada (2021, henceforth XBK2021), who demonstrated prosodic adaptation across one type of construction: “It’s X-ing” (e.g., “It’s cooking”). XBK2021 resynthesized 11-step continua (Figure 1, for details see Supplementary Information) between naturally produced falling (H\* L-L%) and rising (H\* L-H%) terminal pitch. The acoustic continua allowed for testing of listeners’ categorization judgments of items into question vs. statement prosody. This was analogous to creating acoustic continuum between contrasting phonemes (e.g., with an unambiguous instance of “/p/each” at one end, and an unambiguous instance of “/b/each” at the other, see Norris et al. 2003; Clayards et al., 2008).



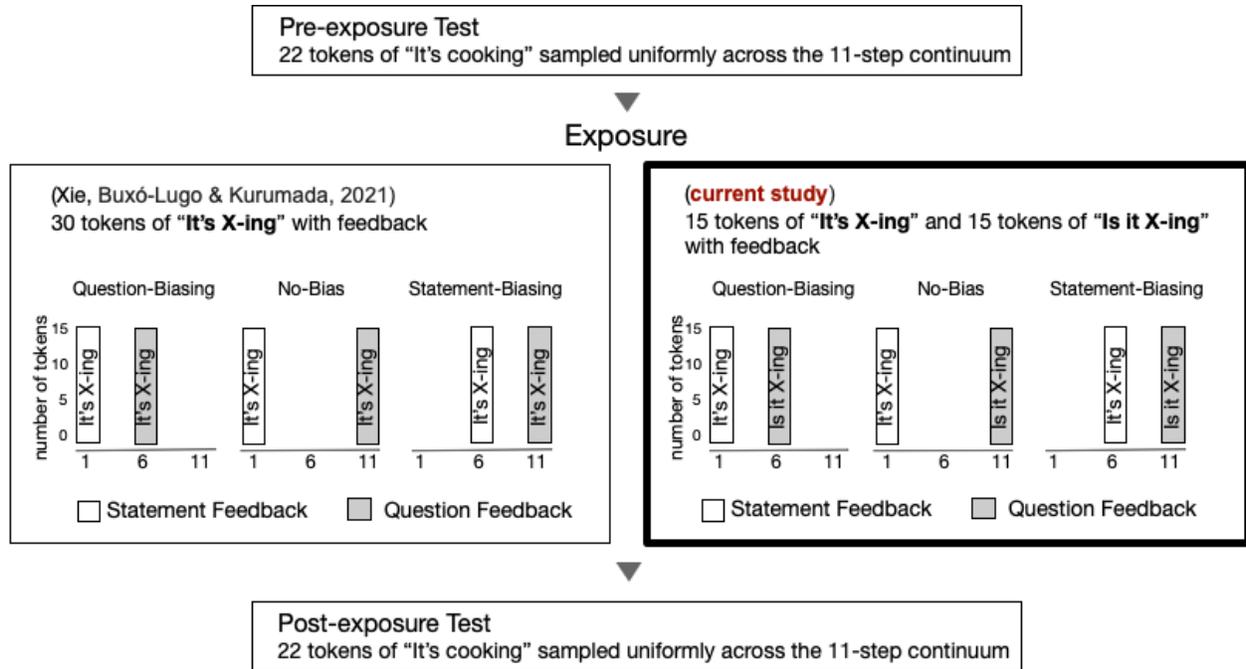
**Figure 1.** Pitch track and durational measures of test items (“It’s cooking”) used both in XBK2021 and in the current study (reprinted from XBK2021).

To create exposure stimuli, we recorded the same female native speaker of American English. The speaker produced the construction with the interrogative syntax “Is it X-ing” with the same five exposure verb types used in XBK2021 (i.e., booting, cooling, losing, moving, muting). These were spoken with the falling terminal pitch to form a basis of the same resynthesis procedure. The recordings were divided into three regions (i.e., is it | X | ing), and we imposed the f0 and duration measurements used in XBK2021 (taken from “It’s moving”) onto these new items to create 11 steps with Step 6 as the middle (stimuli available here: <https://osf.io/hdfk/>).

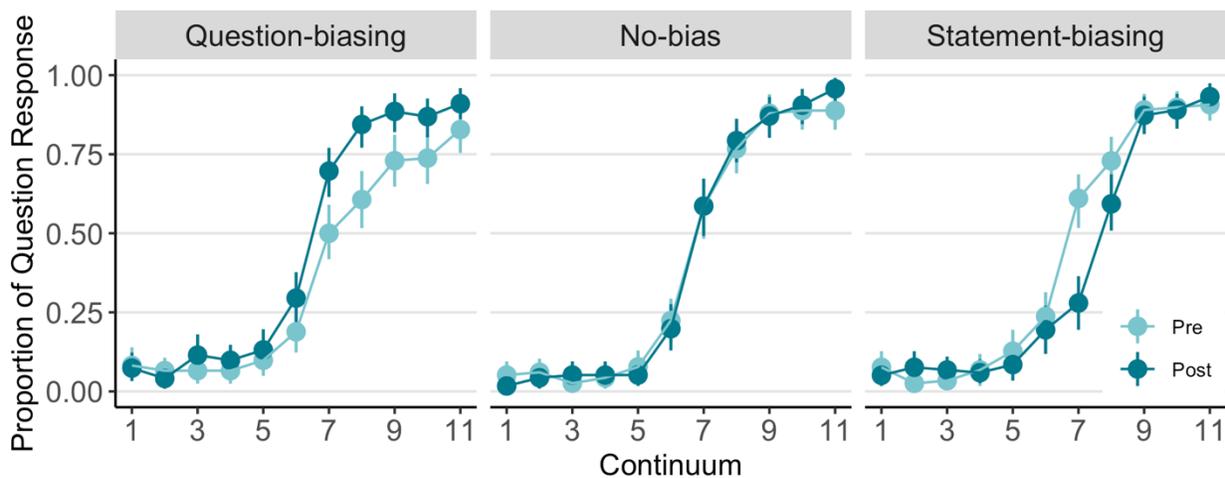
### 2.3 Procedure

The procedure was identical to XBK2021 except for the combinations of tokens heard during the exposure phase (Figure 2). Throughout the experiment, participants were asked to judge whether a given utterance was meant to be a question or a statement and provide their answer in a two-alternative forced choice task (2AFC) by clicking one of the two toggleable buttons labeled as “Question” or “Statement.” During the pre- and post-test, as in XBK2021, participants heard 22 tokens of “It’s cooking” sampled uniformly along the 11-step continuum in a randomized order, with no feedback.

In exposure, participants were randomly assigned to one of the three conditions: *question-biasing*, *no-bias*, or *statement-biasing*. In all the conditions, they continued to answer 2AFC questions, now receiving feedback. In the no-bias condition, exposure tokens with statement feedback were sampled from Step 1 (the lowest pitch rise), and those with question feedback from Step 11 (the highest pitch rise) (Figure 1). In the statement-biasing condition, exposure tokens with statement feedback were sampled from Step 6, and those with question feedback from Step 11. In the question-biasing condition, exposure tokens with statement feedback were sampled from Step 1, and those with question feedback from Step 6. All participants heard the five exposure token types six times each: three times with statement feedback and three times with question feedback (30 total trials).



**Figure 2.** Schematic for the flow of the experiment in XBK2021 (left) and in the current study (right). In both, listeners repeatedly categorized tokens sampled uniformly from the 11-step continua in the pre- vs. post-exposure test blocks (22 tokens each). The only difference is the syntactic construction of a subset of the tokens used in the exposure phase. In the current study, all the exposure tokens associated with question feedback (gray bars) were with the interrogative syntax (“Is it X-ing”). Note that the interrogative syntax was used for tokens from either Step 6 or Step 11. It was never paired with the falling contour (i.e., Step 1).



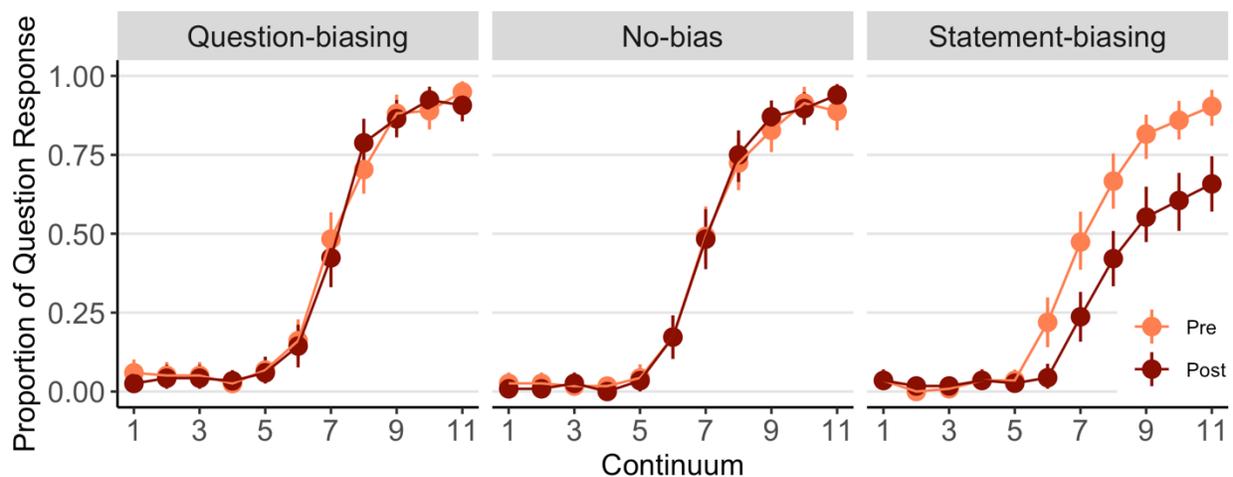
**Figure 3.** Proportions of question responses in the pre- vs. post-exposure test blocks across conditions in XBK2021. Error bars indicate bootstrapped 95% confidence intervals.

The sole difference between XBK2021 and the current study is that exposure tokens with question feedback were hosted in the interrogative syntax (“Is it X-ing”) rather than in the declarative syntax (“It’s X-ing”) (Figure 2). Note that the interrogative syntax was used for Step 6 (i.e., ambiguous between falling and rising) or Step 11 (rising) depending on the condition but *never* for Step 1 (falling). If listeners were tracking pitch contours irrespective of the syntactic contexts, prosodic information to be learned from the exposure stimuli would be identical across the two studies, predicting replication of the results from XBK2021. Alternatively, if adaptation is conditioned on the syntactic context, it was expected only in the statement-biasing condition, where the ambiguous exposure tokens and test tokens shared the syntactic context. Limited or no adaptation was predicted for the question-biasing condition because the ambiguous tokens were heard with the different, interrogative syntax.

In both scenarios, the presence of contrasting syntactic contexts may prompt listeners to make an additional form-based inference (Grice, 1975) along the lines of “If the talker intended to ask a question, she should have used the interrogative syntax. ‘It’s X-ing’ should therefore signal the statement meaning.” Such an inference, although important in understanding pragmatic processing, is orthogonal to the prosodic adaptation we investigate here. We addressed this possibility by assessing the results of the no-bias condition. The potential form-based inference, if present, would trigger an overall bias towards statement responses at post-test.

### 3 Results

Results are summarized in Figure 4. To assess the effect of exposure for each of the conditions, we constructed a logistic mixed effects model using the R package BRMS (Bayesian Regression Models using Stan, Bürkner, 2019). The model had question responses as a dependent variable (coded as 1 = question or 0 = statement) and included block (post- vs pre-exposure, dummy-coded), condition (dummy-coded, with the no-bias condition as the comparison group), their interactions, and continuum step (centered). The random effects structure contained random slopes and intercepts per subject for block, condition, and continuum. Details of the model can be found in Supplementary Information.



**Figure 4.** Proportions of question responses in the pre- vs. post-exposure test blocks across conditions in the current experiment. Error bars indicate bootstrapped 95% confidence intervals.

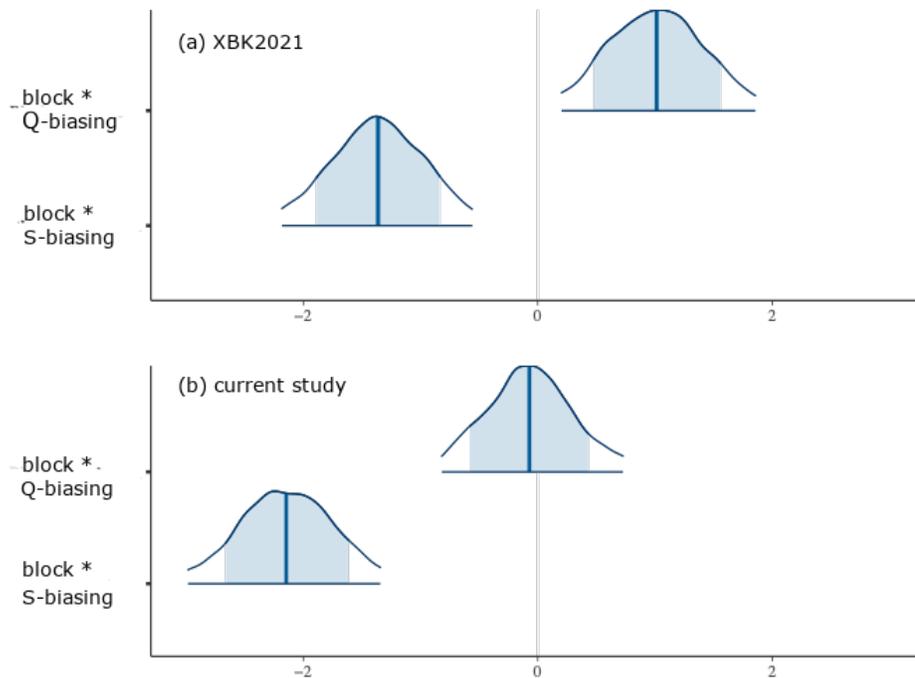
A model summary is provided in Table 1. BRMS, unlike frequentist models, estimates the posterior probability distribution given the observed data and specified priors. We report 95% credible intervals (CIs) to evaluate whether a given coefficient is meaningfully different from zero. As expected, the model found evidence for a main effect of the continuum step: Participants were more likely to provide a question response at higher steps along the continuum ( $\beta = 1.59$ , 95% CI = [1.44, 1.76];  $\Pr(\beta = 0) \approx 0$ , Evidence Ratio = Infinite). Crucially, there was strong evidence for an interaction between the blocks and the statement-biasing condition ( $\beta = -2.15$ , 95% CI = [-2.98, -1.34];  $\Pr(\beta < 0) \approx 1$ , Evidence Ratio = Infinite) while there was virtually no evidence for the interaction between the blocks and the question-biasing condition ( $\beta = -0.07$ , 95% CI = [-0.82, 0.73];  $\Pr(\beta > 0) \approx 0.42$ , Evidence Ratio = 0.73). This difference supports the prediction that listeners condition their prosodic adaptation on syntactic contexts of heard items. In addition, the absence of the main effect of the block eliminated the possibility that the increased statement responses in the statement-biasing condition was simply due to a form-based inference between the two constructions.

**Table 1:** Summary of population-level (fixed) effects of the logistic mixed effects model using BRMS.

	Estimate	Estimated Error	95% CI
Intercept	-1.89	0.24	[-2.38, -1.41]
Block (Post-test vs Pre-test)	0.07	0.28	[-0.49, 0.63]
Question-biasing No-bias	vs. -0.10	0.35	[-0.79, 0.58]
Statement-biasing No-bias	vs. 0.26	0.35	[-0.42, 0.96]
Continuum	1.59	0.08	[1.44, 1.76]
Block * Question-biasing	-0.07	0.40	[-0.82, 0.73]
Block * Statement-biasing	-2.15	0.42	[-2.98, -1.34]

To directly compare the results of the current experiment against those from XBK2021, we conducted a post-hoc analysis over a combined dataset using the same model specifications as described above (Supplementary Information). Figure 5 compares the posterior probability distributions for the critical two-way interactions across the experiments. The posterior distributions of coefficients for the block \* statement-biasing interactions (in log-odds) were reasonably far away from 0 in both experiments; participants were significantly less likely to respond “question” after the exposure phase. In contrast, the effect of block \* question-biasing condition was virtually nonexistent, unlike in XBK2021. Finally, posterior distributions of the three-way interaction between the block \* question-biasing condition \* experiment (XBK2021

vs. Current) was also meaningfully different from zero. This corroborates the idea that adaptation was blocked when the exposure and test tokens were associated with distinct syntactic contexts.



**Figure 5.** Posterior distributions of the two-way interaction terms in (a) XBK2021 and (b) the current study. Q-biasing = question-biasing; S-biasing = statement-biasing. The X-axis represents log-odds. Shaded areas and cut-off points of the density curves indicate 80% and 95% Highest Posterior Density Intervals (HPDIs), respectively.

## 4 Discussion

The importance of talker-normalized pitch and speech rate for prosodic processing is intuitive and has garnered much attention. Here, we examined the hypothesis that listeners’ knowledge of prosodic variability includes fine-grained phonetic variations conditioned on their syntactic contexts. Between Xie et al. (2021) and the current study, the prosodically identical exposure tokens — “It’s X-ing” vs. “Is it X-ing”— yielded strikingly distinct effects on listeners’ post-test categorization judgments. Critically, participants who heard the ambiguous exposure tokens with the interrogative syntax did not alter their judgments on tokens with the declarative syntax before and after the exposure. If adaptation relied solely on adjusting estimates of the talker’s overall base vocal pitch and/or pitch range, this result would not be predictable (for novel, corroborating evidence that prosodic adaptation requires a linguistically intelligible input beyond a talker’s pitch information, see Bosker (2021)).

The absence of adaptation in the question-biasing condition and the (numerically) enhanced adaptation in the statement-biasing condition can reveal interesting dynamics present in speech adaptation. One possibility is that the patterns could be indicative of error-based learning as a necessary component of speech adaptation (e.g., Dell & Chang, 2013; Fine &

Jaeger, 2013). A prosodically ambiguous exposure token in the question-biasing condition was consistently marked as a question via the interrogative syntax (i.e., the subject-auxiliary inversion). Listeners could therefore rely solely on the syntactic marking to answer the 2AFC questions, which might make prosodic adaptation less consequential for comprehension. Although possible, this account runs counter to prevailing theories of speech perception, where recalibration of speech categories is considered largely implicit and automatic. For example, exemplar-based accounts assume that each new exemplar is indiscriminately stored. Adaptation results from subsequent input being categorized with reference to all previously experienced exemplars (e.g., Hay & Drager, 2007; Johnson, 2006; Pierrehumbert, 2001). We know of no account that indicates that speech adaptation occurs exclusively when it is consequential for lexical/meaning disambiguation. If anything, labeling (or disambiguating) information from a lexical or syntactic context (e.g., “croco[?]ile” for the /d/ category) is thought to guide phonetic adaptation rather than block it (e.g., Norris et al., (2003); Kraljic & Samuel (2005)). For this reason, the availability of the syntactic marking of question-hood alone may be insufficient to account for the discrepancy between the current results and those from XBK2021.

An alternative, though not mutually exclusive, explanation would involve the role of listeners’ fine-grained, phonetic expectations about realizations of declarative vs. interrogative questions, against which the current exposure input was processed. As mentioned in the introduction, declarative questions in English tend to show an overall more pronounced level of terminal pitch rise than interrogative questions (Haan, 2001; Grabe, 2002). Note that the acoustic continua used in the current study were created originally from values taken from declarative question. This could mean that a token of the interrogative question that was meant to be prosodically “ambiguous” (Step 6) may have had a terminal pitch rise large enough for an interrogative question. If so, listeners in the question-biasing condition might not have had much evidence that the talker’s uses of prosody were in any way deviant from what is normally expected, and hence no adaptation was needed. According to the same logic, in the statement-biasing condition, the pitch rise experienced with the interrogative syntax (Step 11) during exposure might have been more extreme than what would be generally expected. Participants might have then inferred that post-exposure test tokens must have an even more pronounced rise for them to count as (declarative) questions. This correctly predicts the significant adaptation seen in the statement-biasing condition.

An empirical test of this hypothesis requires assessments of natural production data, detailing whether and how much prosodic variability is observed for different syntactic contexts (e.g., declarative vs. interrogative questions). Such production data can be used to approximate listeners’ expectations at the outset of an experiment. Once made available, a large-scale corpus of this kind will help address other outstanding questions. Chief among them is which syntactic constructions serve as conditioning contexts for prosodic adaptation. Do listeners store all possible syntax-prosody combinations? Or do they represent only those that *typically* vary within/across talkers? Recent development of computational approaches has made it possible to quantify the amount of variability associated with a given context, be it linguistic (e.g., lexico-syntactic) or socio-indexical (Chodroff & Wilson, 2019; Kleinschmidt, 2019). One can estimate the amount of perceptual benefit a listener could, in principle, gain if the variability is effectively conditioned on context (e.g., How much more accurately can one distinguish a question from a statement if the identity of a talker is known vs. unknown? Xie, Buxó-Lugo & Kurumada, 2021). Extending these approaches will facilitate investigations into how listeners may represent and weight multiple sources of variability to achieve reliable comprehension of speech prosody.

In summary, listeners tune into phonetic details of prosody as they adapt to individual talkers. This finding resonates with existing linguistic and psycholinguistic theories, wherein listeners are believed to leverage prosodic details to distinguish between subtle shades of meaning. What we called declarative questions in this study, for instance, can be further subdivided according to their form and meaning (Jeong & Potts, 2016). However, few accounts have been given so far to explain how listeners detect these subtle differences in the sea of talker-variability in human speech. The current finding suggests that listeners can achieve this by flexibly adapting their prosodic processing in a manner attuned to variations expected across linguistic (e.g., syntactic) contexts. This way of reasoning is analogous to position-dependent adaptation of phonemes (i.e., word initial, medial, and final). Overlaying variations stemming from talkers and talker groups can be best learned if the perceived stimulus features are normalized against the *a priori* expectations given a position effect. We conclude that, also in prosodic processing, listeners may condition their expectations according to lexical and syntactic contexts, creating a basis on which to accommodate cross-talker variability.

## Open Practices Statement

The data, stimuli, and supplementary analyses for the experiment are available at <https://osf.io/hdfk/>. The experiment was not pre-registered.

## References

- Ades, A. (1974). How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, *16*, 61–66.
- Arvaniti, A. (2011). The representation of intonation. In *Blackwell Companion to Phonology* (pp. 757–780). Oxford, UK: Oxford University Press.
- Beckman, M. E. (1995) Local shapes and global trends. *Proceedings International Conference of Phonetic Sciences*, Stockholm, Vol. II, (pp. 100-107).
- Bishop, J., & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex, *The Journal of the Acoustical Society of America* 2012, *132* (2), 1100-1112.
- Bolinger, D. (1964). Intonation as a universal. In *Proceedings at Linguistics IX* (pp. 833–844). The Hague: Mouton.
- Bosker, R. H. (2021). Evidence for selective adaptation and recalibration in the perception of lexical stress. *Language and Speech*.
- Breen, M., Kurumada, C., Wagner, M., Watson, D., & Yu, K. (2018). Introducing prosodic variability. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 5. DOI: <http://doi.org/10.5334/labphon.142>
- Bürkner, P.-C. (2019). Bayesian item response modelling in R with BRMS and Stan. Retrieved from <https://arxiv.org/abs/1905.09501>
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.
- Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, *30*(1–2),

1–31.

- Cutler, A. (2015). *Native listening: Language experience and the recognition of spoken words*. MIT Press.
- Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(5), 441–452. <https://doi.org/10.1002/wcs.1355>
- Dahan D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *J Exp Psychol Hum Percept Perform*. 36(3): 704–28. doi: 10.1037/a0017449. PMID: 20515199.
- Dell, G. S., & Chang, F. (2013). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1634), 20120394. <https://doi.org/10.1098/rstb.2012.0394>
- D’Imperio, M., Cangemi, F., & Grice, M. (2016). Introducing advancing prosodic transcription. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1). <https://doi.org/10.5334/labphon.32>
- Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664–1670.
- Fine, A.B., Jaeger T.F., (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37(3):578–91. Doi: 10.1111/cogs.12022.
- Grabe, E. (2002). Variation adds to prosodic typology. *Speech Prosody 2002*, 127–132.
- Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*, vol.3 (pp. 41–58). New York: Academic Press.
- Gunlogson, C. (2003). *True to form: Rising and falling declaratives as questions in English*. New York: Routledge.
- Gussenhoven, C., & Rietveld, T. (1996). On the speaker-dependence of the perceived prominence of F0 peaks. *Journal of Phonetics*, 26, 371–380.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31. 373–405. 10.1016/j.wocn.2003.09.006.
- Haan, J. (2001). *Speaking of questions: An exploration of Dutch question intonation*. LOT Dissertation Series (Vol. 52).
- Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology*, 36(1), 89–103. <https://doi.org/10.1146/annurev.anthro.34.081804.120633>
- Hedberg, N., Sosa, J. M., & Görgülü, E. (2017). The meaning of intonation in yes-no questions in American English: A corpus study. *Corpus Linguistics and Linguistic Theory*, 13(2), 321–368.
- Ito, K., Turnbull, R., & Speer, S. R. (2017). Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses in museum visitors. *Laboratory Phonology*, 8(1). <https://doi.org/http://doi.org/10.5334/labphon.86>
- Jeong, S., & Potts, C. (2016). Intonational sentence-type conventions for perlocutionary effects: An experimental investigation. *Semantics and Linguistic Theory*, 26, 1. <https://doi.org/10.3765/salt.v26i0.3787>
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485–499.
- Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience*, 34(1), 43–68. <https://doi.org/10.1080/23273798.2018.1500698>
- Kleinschmidt, D. F., Weatherholtz, K., & Florian Jaeger, T. (2018). Sociolinguistic perception as

- inference under uncertainty. *Topics in Cognitive Science*, 10(4), 818–834. <https://doi.org/10.1111/tops.12331>
- Kurumada, C., Brown, M., & Tanenhaus, M. K. (2017). Effects of distributional information on categorization of prosodic contours. *Psychological Bulletin and Review*. <https://doi.org/10.3758/s13423-017-1332-6>
- Ladd, D. R. (2008). *Intonational phonology* (2<sup>nd</sup> ed.). Cambridge University Press.
- Lee, C.-Y. (2009). Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study. *Journal of the Acoustical Society of America*, 125, 1125–1137.
- Nakamura, C., Harris, J., & Jun, S.-A. (2019). Listeners' beliefs about the speaker and adaptation to the deviant use of prosody. In *Poster presented at the 32<sup>nd</sup> annual CUNY Conference on Human Sentence Processing, Boulder, CO*.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238.
- Patel, R., & Brayton, J. T. (2009). Identifying prosodic contrasts in utterances produced by 4, 7, and 11 year old children. *Journal of Speech, Language, and Hearing Research*, (June), 790–801.
- Patel, R., Niziolek, C., Reilly, K., & Guenther, F. H. (2011). Prosodic adaptations to pitch perturbation in running speech. *Journal of speech, language, and hearing research*, 54(4), 1051–1059. [https://doi.org/10.1044/1092-4388\(2010/10-0162\)](https://doi.org/10.1044/1092-4388(2010/10-0162))
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *In Frequency and the Emergence of Linguistic Structure* (pp. 137–157). John Benjamins.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311).
- Saindon, M. R., Trehub, S. E., Schellenberg, E. G., & van Lieshout, P. (2017). When is a question a question for children and adults?, *Language Learning and Development*, 13(3), 274-285, DOI: [10.1080/15475441.2016.1252681](https://doi.org/10.1080/15475441.2016.1252681)
- Samuel, A.G. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. *Perception & Psychophysics*, 45(6), 485-493. DOI: 10.3758/BF03208055
- Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, 111, 104070–104070. <https://doi.org/10.1016/j.jml.2019.104070>
- Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G. (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalized storage of intonation. *Speech Communication*, 66, 65-81.
- Tang, C., Hamilton, L. S., & Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. *Science*, 356, 797–801.
- Woodard, K., Plate, R.C., Morningstar, M., Wood, A., & Pollak, S. D., (2021). Categorization of vocal emotion cues depends on distributions of input. *Affective Science* (2021). <https://doi.org/10.1007/s42761-021-00038-w>
- Xie, X., Buxó-Lugo, A., & Kurumada, C. (2021). Encoding and decoding of meaning through structured variability in speech prosody. *Cognition*, 221. <https://doi.org/10.1016/j.cognition.2021.104619>

## Supplementary Information

### S.1 Statistical Analysis of the Data

To analyze the data, we used Bayesian hierarchical models using the Stan (BRMS) modelling language (Carpenter et al., 2017) and the R package *brms* (Bürkner, 2019). All data tables and R scripts are available here: <https://osf.io/hdftk/>. Here, we describe the statistical analyses we conducted and assumptions applied in the analyses. We acknowledge that we are in debt to careful documentations provided in prior studies, especially Roettger and Baer-Henney (2018). We chose BRMS over a canonical frequentist model for the current analysis for two reasons.

First, within the constraints of the current design, the Bayesian framework allowed us to construct a model with the maximal random effect structure justified by the data. We have constructed linear mixed effects models using the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015) with the random intercepts and slopes (as recommended by Schielzeth & Forstmeier (2008); Barr, Levy, Scheepers, & Tily (2013); Bates, Kliegl, Vasishth, & Baayen (2015)), but those models failed to converge. BRMS allowed us to fit the maximal random effect structure, which is considered more conservative.

Second, the Bayesian framework enables us to quantitatively estimate the likelihood of a particular hypothesis given the observed data. In other words, this modeling method can quantify our uncertainty about the parameters of interest, “which frees us from committing to hard cut-off points for statistical significance (such as the arbitrary .05 alpha level)” (Roettger & Baer-Henney, 2018). This was better suited than a frequentist model to our goal of comparing the results of the two experiments to infer *an extent to which* the current design induced results different from those in XBK2021.

In analyzing the current experiment, we fit hierarchical regression models to log-odds of binomial responses (question vs. statement) predicted by Conditions (question-biasing, no-bias, statement-biasing, dummy coded as the no-bias condition as the baseline), Blocks (pre, post, dummy-coded), their interactions, and Continuum (1-11, centered). The models included a random-effect structure that included by-subject random slopes and intercepts. Following recommendations from the statistical literature (Gelman et al., 2013; Bürkner, 2019), we used the default priors in BRMS, which are meant to be weakly informative. However, it is worth noting that specifying priors based on the results from XBK2021 did not meaningfully change the pattern of results.

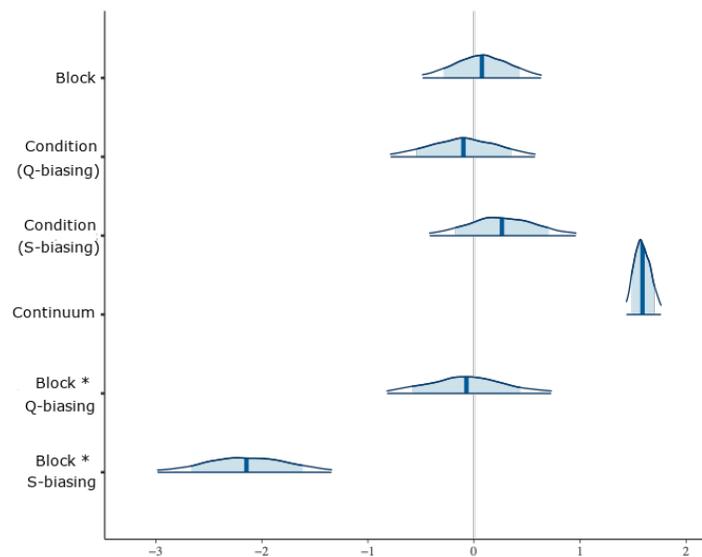
The model was fit via Stan's No-U-Turn sampler — a Hamiltonian Monte Carlo Method with many computationally desirable properties (Monnahan, Thorson, & Branch, 2017). Four sampling chains with 2000 iterations were run for each model, with a warm-up period of 1000 iterations. The chains mixed well (e.g., all R-hats  $\sim 1.001$  or closer to 1), and there were no divergent transitions after warm-up.

We report 95% credible intervals (CIs). A 95% credible interval “demarcates the range of values that comprise 95% of the probability mass of our posterior beliefs” (Roettger & Baer-Henney, 2018). It is generally accepted that there is compelling evidence for an effect if 0 is not included in the 95% CI. As part of the model summary, we report an evidence ratio as derived by the “hypothesis” function included in BRMS, as well as an estimated probability of the draws

from the posterior distributions of the critical interactions to be larger or smaller than 0 in an expected direction, that is  $\Pr(\beta < 0)$  or  $\Pr(\beta > 0)$ . This estimate provides supportive evidence for the alternative hypothesis, as opposed to the null hypothesis, if it is close to 1.

The model output is summarized in Table 1 in the manuscript. Here we include a corresponding visual representation of posterior uncertainty distributions for the fixed effects in Figure s1. As we mentioned above, a given fixed effect can be considered “significant” (meaningfully distinguishing the alternative from the null hypothesis) if a distribution is away from 0 by a sufficient margin. As can be seen in Figure s1 above, the distributions for the main effects of Block and Conditions (question-biasing / statement-biasing) cross 0, hence those effects are unlikely to be significant. On the other hand, the main effect of Continuum is, as expected, highly likely to be significant. That is, participants were more likely to provide the question response for the items sampled from a region closer to the higher end of the continuum.

Distributions of the two-way interaction terms, Block \* question-biasing / statement-biasing Conditions, exhibited different patterns. The interaction between Block and statement-biasing condition is significantly away from 0, suggesting that participants in the statement-biasing condition were less likely to provide the question response in the post-exposure phase. In contrast, the interaction between Block and question-biasing condition crosses 0. This is a pattern expected under the prediction that intonation adaptation is conditioned on syntactic constructions: In the question-biasing condition where the question meaning was expressed with distinct constructions between the exposure and test stimuli phases, the adaptation was unlikely to occur.



**Figure s1.** Posterior uncertainty distributions of the fixed effects of the BRMS model. The X-axis represents log-odds. S-biasing = Statement-biasing Condition; Q-biasing = Question-biasing Condition. Shaded areas indicate 80% Highest Posterior Density Intervals (HPDIs), and the distributions are cut off at 95% HPDIs.

## S.2 Statistical Analysis of the Data from XBK 2021

The data from XBK2021 were analyzed with the same exact model structure and coding schemes as those described for the current study. Packages and tools used were identical to those in the

current study to allow for comparisons of the model outcomes from the two studies. The summary of the fixed effect in the model of the XBK2021 data is presented in Table s1 below. Overall, the results were similar to those from the present study. A key difference was that, whereas the present study found no evidence for the presence of a Block \* question-biasing interaction, XBK2021 did find evidence for this interaction in the expected direction ( $\beta = 1.01$ , 95%CI = [.20, 1.86];  $\Pr(\beta > 0) = 1$ ). Participants in the question-biasing condition in XBK2021, unlike those in the current experiment, were more likely to provide the question response during the post-exposure block.

Table s1. Summary of population-level (fixed) effects of the logistic mixed effects model using BRMS run on the combined dataset from XBK2021 and the current study.

	Estimate	Estimated Error	95% CI
Intercept	-1.62	0.28	[-2.18, -1.08]
Block (Post-exposure vs Pre-exposure)	0.06	0.30	[-0.53, 0.65]
Question-Biasing vs No-bias	-0.35	0.41	[-1.15, 0.45]
Statement-Biasing vs No-bias	0.17	0.38	[-0.58, 0.94]
Continuum	1.70	0.11	[1.50, 1.91]
Block * Question-Biasing	1.01	0.42	<b>[0.20, 1.86]</b>
Block * Statement-Biasing	-1.36	0.41	[-2.18, -0.56]

### S.3 Statistical Analysis of the Combined Dataset

To directly compare the results of the two experiments, we conducted an analysis on a combined dataset including the data from both XBK2021 and the current experiment. The model specifications are identical to what we described under the Results section of the manuscript and Section 1 above except for the Experiment factor (XBK2021 vs. Current, dummy-coded). This was included as a fixed effect with an interaction term with Block and Condition, and in the random effects structure as independent slopes and intercepts per subject. By including the interaction between Experiment and the crucial Block \* Condition interaction, we can test whether the effects of Block \* Condition were significantly different between XBK2021 and the current experiment.

The summary of the fixed effects of the model is given in Table s2 below. We found evidence for a three-way interaction between Block, Condition, and Experiment for the question-biasing conditions ( $\beta = -1.18$ , 95%CI = [-1.84, -0.54];  $\Pr(\beta < 0) = 1$ ). Participants in the current experiment, as compared to those in XBK2021, were significantly *less* likely to provide the question response after the Exposure phase. This supports the conclusion that the degrees of adaptation seen in the question-biasing condition differed across these experiments.

We note also that the three-way interaction was significant for the S-biasing ( $\beta = -0.85$ , 95%CI = [-1.47, -0.23];  $\Pr(\beta < 0) = 1$ ). This likely reflects the strong bias we observed at the higher region along the continuum (Figure 4). Compared to those in XBK2021, participants in the statement-biasing condition in the current experiment were more likely to provide the

statement response in the post-exposure test. As we discuss in the Discussion, this could be due to the fact that the exposure tokens in this condition strongly biased the listener to expect a prominent rise for a declarative question. In Exposure, listeners in this condition heard “It’s X-ing” tokens from Step 6 as associated with the statement meaning and “Is it X-ing” tokens from Step 11. This pattern of exposure supports the following observations:

- 1) The “It’s-Xing” tokens were associated with more terminal rise than expected.
- 2) The tokens of “Is it X-ing” have a more prominent terminal pitch rise than would be generally expected in productions. (Recall that these tokens were resynthesized based on the intonational features of a declarative question.)

Based on these observations, listeners might have concluded that this particular talker tends to produce more extreme terminal pitch rise than normally expected. This would support an inference that a *declarative* question produced by the same talker would have even more extreme rise. If this is the case, we can straightforwardly explain the increased statement responses in the post-exposure test. Though the validity of such inferences needs to be directly verified in future studies, we conclude here that listeners’ judgments as observed in the post-exposure test are predictable based on the types of intonation variations typically associated with the two distinct constructions used in the experiment.

Table s2. Summary of population-level (fixed) effects of the logistic mixed effects model using BRMS run on the combined dataset from XBK2021 and the current study.

	Estimate	Estimated Error	95% CI
Intercept	-1.27	0.18	[-1.64, -0.91]
Block (Post-Exposure vs Pre-Exposure)	0.02	0.17	[-0.32, 0.34]
Question-Biasing vs No-bias	-0.39	0.25	[-0.88, 0.11]
Statement-Biasing vs No-bias	0.13	0.25	[-0.36, 0.63]
Experiment (Current vs. XBK2021)	-0.30	0.26	[-0.81, 0.22]
Continuum	1.24	0.05	[1.14, 1.35]
Block * Question-Biasing	0.96	0.23	[0.51, 1.40]
Block * Statement-Biasing	-0.68	0.23	[-1.13, -0.24]
Block * Experiment	-0.02	0.22	[-0.44, 0.40]
Question-Biasing * Experiment	0.54	0.37	[-0.18, 1.28]
Statement-Biasing * Experiment	-0.29	0.38	[-1.03, 0.45]
Block * Question-Biasing * Experiment	-1.18	0.33	[-1.84, -0.54]
Block * Statement-Biasing * Experiment	-0.85	0.32	[-1.47, -0.23]

## References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. Retrieved from <http://arxiv.org/abs/1506.04967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bürkner, P.-C. (2019). Bayesian Item Response Modelling in R with brms and Stan. Retrieved from <https://arxiv.org/abs/1905.09501>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software; Vol 1, Issue 1 (2017)*. Retrieved from <http://dx.doi.org/10.18637/jss.v076.i01>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.
- Monnahan, C. C., Thorson, J. T., & Branch, T. A. (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution*, 8(3), 339–348. <https://doi.org/10.1111/2041-210X.12681>
- Roettger, T. B., & Baer-henney, D. (2018). Toward a replication culture: Speech production research in the classroom. *PsyArXiv*, 1–26. <https://doi.org/10.17605/OSF.IO/9KYWF>
- Schielzeth, H., & Forstmeier, W. (2008). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, 20(2), 416–420. <https://doi.org/10.1093/beheco/arn145>



# A UNIFIED OPERATOR MOVEMENT ANALYSIS FOR ADVERBIAL CLAUSES AND RELATIVE CLAUSES IN SWAHILI

ZHENDONG LIU

*Department of Linguistics, University of Southern California*<sup>1</sup>

## Abstract

This study addresses structural similarities and differences between adverbial clauses and relative clauses in Swahili and proposes a unified operator-movement for both. Its main contributions are twofold: first, it supports numerous theoretical hypotheses with facts from Swahili, including the operator movement analysis of adverbial clauses as well as relative clauses, and a truncated left periphery of adverbial clauses. Second, it challenges some of the previous analyses of Swahili relative clauses and provides a novel account with insights gained from adverbial clauses.

## 1 Introduction

The type of adverbial clause in Swahili analyzed in this paper is introduced by a so-called *subordinator* (-*vyo* and -*po* in the examples below). The kinds of adverbial meaning that are available in this construction are the coincidence of time (1a, 1b), location (1b), and manner (1c)<sup>2</sup>.

---

<sup>1</sup> This paper was completed during my master's degree at the Department of Linguistics, University of Rochester

<sup>2</sup> Abbreviations in this paper:

SG - Singular; SM - Subject Marker; OM - Object Marker; PST - Past Tense; FUT - Future Tense; SUB - Subordinator; REL - Relative Marker; IND - Indicative; APPL - Applicative; PERF - Perfective; DEM - Demonstrative; PASS - Passive; NEG - Negation



relative clause constructions while it is not in adverbial clauses. I will argue that the root cause of the word order asymmetry is the truncated left periphery of adverbial clauses. Unless otherwise noted, all data in this paper come from original fieldwork.

## 2 Adverbial clauses and relative clauses: similarities

Relative clauses in Swahili have characteristic properties. To explore the similarities between Swahili adverbial clauses and relative clauses, we must diagnose whether adverbial clauses also have such properties. A property of Swahili relative clauses is their three different formation strategies recognized in many previous works (Ermisch (2012); Ngonyani (2006); Barrett-Keach (1980); Schadeberg (1989) among others), as shown in (3).

(3) Three relative constructions in Swahili (Ngonyani (2006)).

- a. vi-tabu [<sub>CP</sub> amba-**vyo** Juma a-li-nunu-a] ni ghali  
8-book [<sub>CP</sub> amba-**8REL** Juma 1SM-PST-buy-IND] be expensive  
The books Juma bought are expensive.
- b. vi-tabu [<sub>CP</sub> a-li-**vyo**-nunu-a Juma] ni ghali  
8-book [<sub>CP</sub> 1SM-PST-**8REL**-buy-IND Juma] be expensive  
The books Juma bought are expensive.
- c. vi-tabu [<sub>CP</sub> a-nunu-a-**vyo** Juma] ni ghali  
8-book [<sub>CP</sub> 1SM-buy-IND-**8REL** Juma] be expensive  
The books Juma buys are expensive

In these constructions, the class 8 relative marker (-*vyo*) has three different surface positions. In (3a), -*vyo* attaches to the complementizer *amba*. In (3b), it appears between the tense marker and the verb stem. In (3c), where the tense marker is absent, it is attached at the end of the complex verb after the final vowel. Interestingly, subordinators in adverbial clauses can appear in the exact same three positions, as shown in (4).

- (4) a. ni-li-pig-a chafya [<sub>CP</sub> amba-**po** Baraka a-li-kat-a] ki-tungu]  
1SG-PST-hit-IND sneeze [<sub>CP</sub> amba-**SUB** Baraka 1SM-PST-cut-IND 7-onion]  
I sneezed where Baraka cut the onion
- b. ni-li-pig-a chafya [<sub>CP</sub> Baraka a-li-**po**-kat-a] ki-tungu]  
1SG-PST-hit-IND sneeze [<sub>CP</sub> Baraka 1SM-PST-**SUB**-cut-IND 7-onion]  
I sneezed when Baraka cut the onion
- c. ni-ta-pig-a chafya [<sub>CP</sub> Baraka a-kat-a-**po**] ki-tungu]  
1SG-FUT-hit-IND sneeze [<sub>CP</sub> Baraka 1SM-cut-IND-**SUB** 7-onion]  
I will sneeze when Baraka cuts the onion.

Beside the common positions of the relative marker and the subordinator, another similarity is observed in the morphology of relative markers and subordinators. Notice that both relative markers and subordinators have an *-o* ending. Barrett-Keach (1986) points out that the *-o* ending is characteristic of pronominal clitics in Swahili. As shown in (5), pronominal clitics in Swahili consist of a noun class agreement prefix and the *-o* morpheme.

- (5) **Kata nyama kwa-cho**  
 cut meat with-7o  
 Cut the meat with it  
 (Barrett-Keach (1986))

The relative markers (*-vyo* and *-cho*) and the subordinators (*-vyo* and *-po*) that we have seen so far can also be decomposed in this way. Notice that in the relative clauses in (2) and (3), relative markers have the same noun class feature as the relativized NPs. On the other hand, the subordinators *-vyo* and *-po* have the same forms as the known class 8 and class 16 relative markers respectively. Thus, both relative markers and subordinators can be analyzed as the exponent of a pronominal clitic *-o* with a noun class feature.<sup>3</sup>

The third similarity is that adverbial clauses can actually act as relative clauses, modifying an overt NP denoting time, place and manner (6).

- (6) a. ni-li-pig-a            chafya **mda** [CP Baraka a-li-**po**-kat-a            ki-tungu]  
 1SG-PST-hit-IND sneeze time [CP Baraka 1SM-PST-SUB-cut-IND 7-onion]  
 I sneezed at the time when Baraka cut the onion.
- b. ni-me-li-on-a            gari **mahali** [CP Baraka a-li-**po**-li-on-a  
 1SG-PERF-5OM-see-IND car place [CP Baraka 1SM-PST-SUB-5OM-see-IND  
 jana]  
 yesterday]  
 I have seen the car at the place where Baraka saw it yesterday
- c. ni-ta-m-pend-a            huju    m-bwa **jinsi** [CP amba-**vyo** Baraka  
 1SG-FUT-3OM-like-IND 3-DEM 3-dog manner [CP amba-SUB Baraka  
 a-li-m-pend-a]  
 1SM-PST-3OM-like-IND]  
 I will love this dog the way Baraka used to love him.

Thus, the collected data demonstrate that this type of adverbial clause in Swahili shares most of the recognized characteristic properties of relative clauses, except for the order of the subject and the complex verb in the *amba*-less constructions: In relative clauses, the subjects must appear after the verb, while this order is not obligatory in adverbial clauses. This phenomenon will be accounted for in section 5. For now, I conclude that adverbial clauses of this kind have more similarities with relative clauses in Swahili than previously recognized, and this level of similarity is sufficient to motivate a unified analysis for these two clause types.

<sup>3</sup> The pronominal status of relative markers is also suggested by some early descriptive work which categorized them as relative pronouns (Zwart (1997); Gregersen (1967) among others)

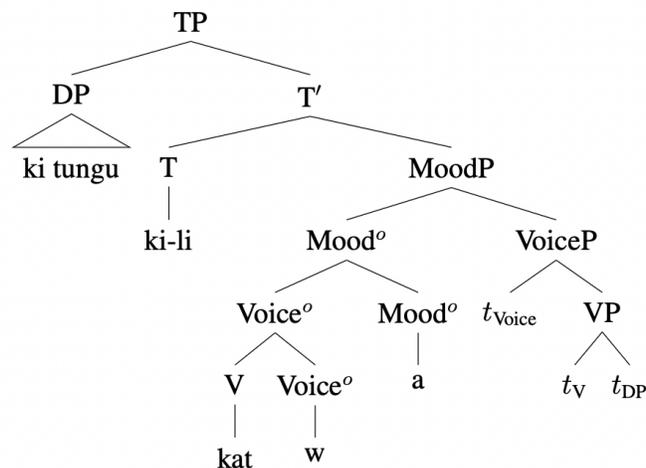
### 3 Proposed analysis for adverbial clauses

In this section, I firstly present an operator movement analysis for the three constructions of adverbial clauses presented in the previous section. I then extend this analysis to relative clauses in section 4. I start with specifying some assumptions about the Swahili clause structure.

For the structure of Swahili verbal morphology, I adopt the analysis of Kilega in Carstens (2005). Like Kilega, The Swahili verb consists of a set of prefixes and suffixes. The prefixes include morphemes for subject agreement and tense, and the suffixes include morphemes indicating mood, passivization and so on. In the case of Kilega, Kinyalolo (2003) argues that the suffixes are attached via V-raising through various functional projections, assuming left head adjunction (Kayne, 1994). This head raising terminates in a position lower than T, which is identified as Mood, and prefixes stay in their base generated positions. Additionally, T is assumed to have an unvalued *Phi*-feature bundled with an EPP feature that agrees with the closest DP in noun class and moves it to the Spec-TP position. Thus, the sentence (7a) has the structure in (7b). In this case, T agrees with the class 7 DP *ki-tungu*, is phonologically realized as *ki-li*, and moves the DP to the Spec,TP position.<sup>4</sup>

- (7) a. ki-tungu ki-li-kat-w-a  
 7-onion 7SM-PST-cut-PASS-IND  
 the onion was cut

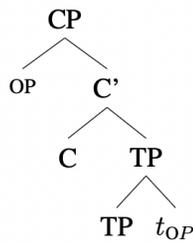
b.



For my analysis, I follow the hypothesis that adverbial clauses are derived by moving an operator to the left periphery (Geis (1970) among others). Thus, another assumption to specify is the extraction position of the operator. Since expressions of time, location and manner of events are syntactically realized as TP or VP adjuncts, it's reasonable to assume that the extraction positions of operators in these adverbial clauses are TP or VP adjunct positions. I only demonstrate my analysis on time and location adverbial clauses, and I assume that the extraction position of the operator in this case is the TP adjunct position. Thus, I assume the following abstract structure for temporal adverbial clauses.

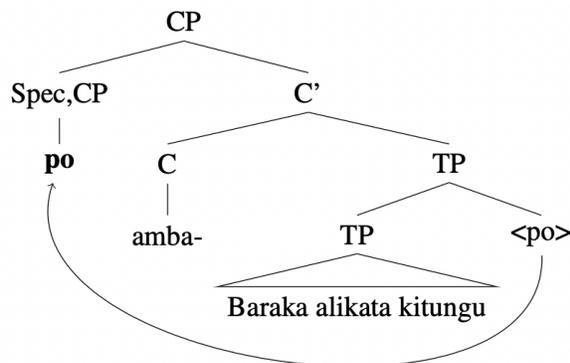
<sup>4</sup> Movement is indicated by trace

## (8) Abstract structure for temporal adverbial clauses



With necessary assumptions specified, I propose that in Swahili adverbial clauses, the subordinator is the exponent of the operator which is base generated with a noun class feature. This proposal is motivated by the fact that in the operator movement analysis for English relative clauses and adverbial clauses, the operators in both cases are *wh*-words, and *wh*-words are a type of pronoun. In Swahili adverbial clauses and relative clauses, the pronominal elements that have the similar *-o* form are the subordinator and the relative marker, as discussed in the previous section. Thus, it is reasonable to analyze them as operators. The operator is then raised to Spec,CP by C. The analysis so far is illustrated in (9).

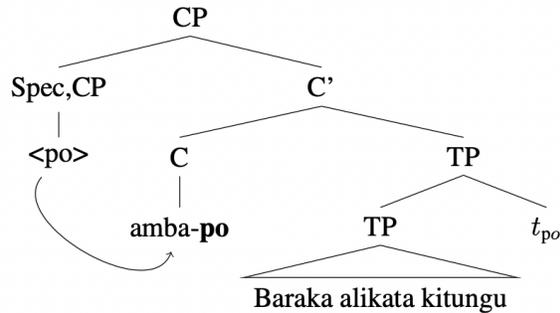
- (9) a. ...[<sub>CP</sub> amba-**po** Baraka a-li-kat-a ki-tungu]  
 ...[<sub>CP</sub> amba-**SUB** Baraka 1SM-PST-cut-IND 7-onion]  
 ... Where baraka cut the onion.
- b. The operator *-o* generated with a class 16 feature, realized as *po*.



Notice that the word order in (9b) is not yet the surface order in (9a), and I proposed that the three surface positions of the operator is derived by post-syntactic movement. Specifically, after all movements in syntax, the operator cliticizes onto the right of some head (C, T or Mood) through m-merger (Matushansky, 2006). I propose that there is a phonological constraint on the destination of cliticization, where the head that the operator cliticizes onto should have more than one syllable. In other words, the operator always cliticizes onto the highest head that is neither empty nor monosyllabic, skipping any empty or monosyllabic heads. Thus, in (9b), *po* cliticizes onto C, since it is the highest head with more than one syllable.

- (10) a. ...<sub>[CP amba-**po** Baraka a-li-kat-a ki-tungu]</sub>  
 ...<sub>[CP amba-**SUB** Baraka 1SM-PST-cut-IND 7-onion]</sub>  
 ... Where baraka cut the onion.

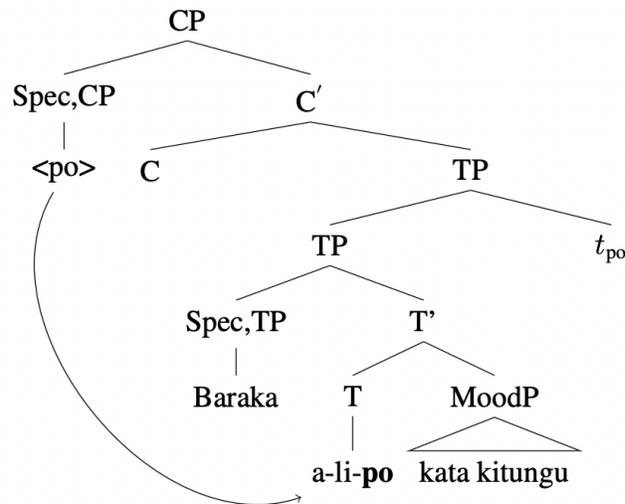
b. The operator *po* cliticizes onto *amba-*



For the second construction without *amba* (4b), since C is phonologically empty, the next available head for the cliticization is T, which is realized with two syllables as *a-li* (11) (Recall that there is no V to T movement in Swahili).

- (11) a. ...<sub>[CP Baraka a-li-**po**-kat-a ki-tungu]</sub>  
 ...<sub>[CP Baraka 1SM-PST-**SUB**-cut-IND 7-onion]</sub>  
 ... where Baraka cut the onion

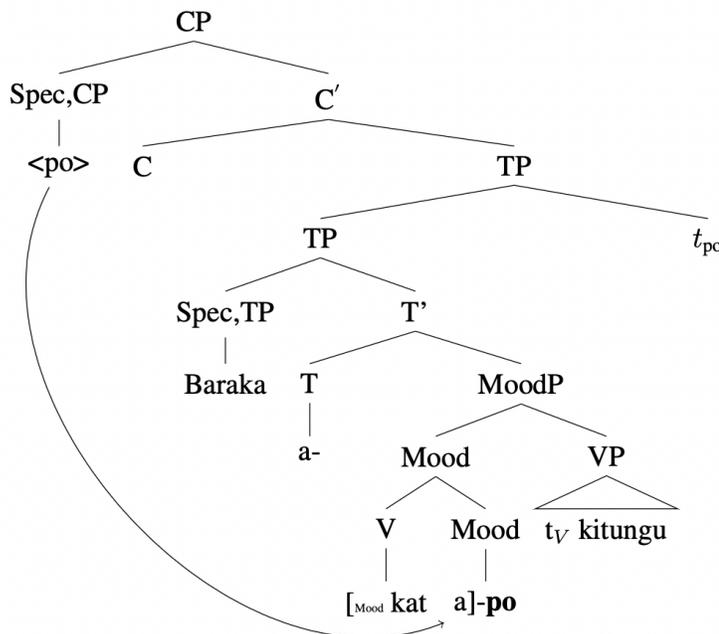
b. The operator *po* cliticizes onto T.



Finally, in the tenseless case (4c), C is empty and T is monosyllabic. For this reason, C and T are not possible destinations for cliticization, and the operator cliticizes all the way onto Mood.

- (12) a. ...[<sub>CP</sub> Baraka a-kat-a-po ki-tungu]  
 ...[<sub>CP</sub> Baraka 1SM-cut-IND-SUB 7-onion]  
 ... When Baraka cuts the onion

- b. The operator *po* cliticizes onto Mood



In summary, my analysis for adverbial clauses proposes that the subordinator is in fact base generated as an operator with a  $\Phi$ -feature of a noun class, and raised to Spec,CP. This operator is then post-syntactically cliticizes onto the highest head pronounced with more than one syllables. However, it is unclear why the noun class features of the operators are class 8 and 16 in these cases. A partial explanation is that class 16 is the noun class for *place*, thus the class 16 operator is used in locative adverbial clauses (Ermisch (2012); Ashton (1947); Mpiranya (2014)). But it is not clear why the class 16 operator can also be used in temporal adverbial clauses, and the class 8 subordinator can be used in both temporal and manner ones. A possible explanation is that the class 16 operator can be alternatively interpreted as *time* in semantics, depending on the context, and similarly the class 8 operator can be interpreted as both *time* and *manner* in appropriate contexts.

## 4 A unified analysis for adverbial and relative clauses

In this section, I will show that the analysis proposed in the previous section can also apply to relative clauses. I argue that relative markers should receive the same analysis as subordinators in adverbial clauses that they are relative operators, and thus the operator movement analysis of relative clauses instead of the raising analysis should be adopted.

### 4.1 Previous analyses of Swahili relative clauses

Some previous analyses of relative clauses in Swahili propose that the relative marker is the exponent of an agreeing relative C, which has an unvalued  $\Phi$ -feature that agrees with the  $\Phi$ -feature



and raises it to Spec,CP. C with the class 7 feature is then pronounced as *cho*. Finally, the complex C *alicho* is lowered to attach to the verb stem, forming the complex verb *alichokata*<sup>5</sup>.

Ngonyani (2001) provides some facts as supporting evidence for a head raising analysis. These facts show that the head of the relative clause is connected to a lower position within the clause, suggesting that the head is base generated under the relative clause. Example (15) is a case where the possessive pronoun *chake* (his/her) inside the head of the relative clause is bound by the quantified noun phrase (QNP) *kila mwandishi* (every writer), while the quantifier is generally assumed to take scope over its c-command domain. Thus, for the QNP to take scope over the head, the head must be base generated below the QNP and be reconstructed to that position at the LF level.

- (15) ki-tabu **ch-ake**<sub>i</sub>      ch-a      kwanza [amba-cho **kila mw-andishi**<sub>i</sub> hu-ji-vun-i-a  
 7-book 7-3SG.POSS 7-CON first      [amba-7REL every 3-writer      HAB-RFL-proud  
 ] hu-w-a      ki-zuri sana  
 ] HAB-be-IND 7-good very  
 Her/his first book for which the writer is very proud is usually very good  
 (Ngonyani, 2001)

Furthermore, example (16) shows that when the nominal part of an idiomatic expression is relativized, the idiomatic meaning is preserved. In Swahili, the expression "*to hit water*" in (16a) means "*to drink alcohol*". (16b) shows that when the nominal component of the idiom, *ma-ji* (water), is the relative head, with a gap inside of the idiom, the idiomatic meaning is preserved. Since idiomatic interpretation is carried by constituents, this fact also suggests that the head noun in (16b) is reconstructed to its base generated position within the relative clause, as the complement of the verb.

- (16) a. komba      a-li-pig-a      ma-ji      sana.  
 1.bushbaby 1SM-PST-hit-IND 6-water much  
 Bushbaby drank much beer.
- b. ma-ji      amba-yo      komba      a-li-ya-pig-a      ya-li-ku-w-a  
 6-water amba-6REL 1.bushbaby 1SM-PST-6OM-hit-IND 6SM-PST-INF-be-IND  
 ma-kali.  
 6-fierce  
 The beer that Bushbaby drank was very strong.  
 (Ngonyani, 2001)

These facts serve as strong evidence for the head raising analysis of relative clauses. Since the operator movement analysis proposed in this work assumes an externally generated head of the relative clause, it is not able to account for these connectivity effects. However, there is also evidence that challenges the head raising analysis, which I will present below. The existence of this contradictory evidence might point to the conclusion that Swahili allows two different derivations of relative clauses — a raising derivation (responsible to the connectivity effects seen above) and

<sup>5</sup> See Ngonyani (2006) for detailed analyses of the other two relative constructions



their own  $\Phi$ -features, they can still have the same semantic denotations as the relativized NPs and be coindexed with them, while having different  $\Phi$ -features.

Finally, the operator-movement analysis of relative clauses is compatible with fact that the relative marker can't co-occur with conditional markers, as shown in (13). It is entirely possible that conditional marker can be alternatively analyzed as operator. In fact, it has been argued in Bhatt and Pancheva (2017) that conditional clauses likely involve clause-internal operator-movement to Spec,CP. Thus the ungrammaticality of (13) likely stems from two operators competing for the Spec,CP position. In other words, the complementary distribution of the two markers is not because they are both complementizers but rather because they are both operators.

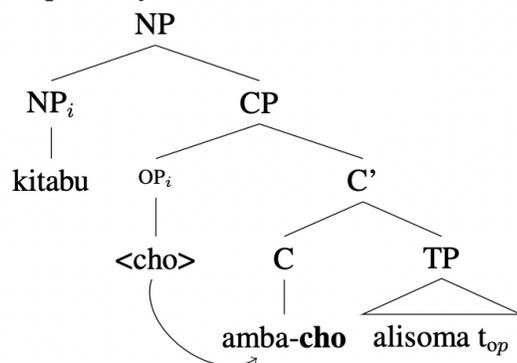
### 4.3 An operator movement analysis for Swahili relative clauses

Having established the operator status of relative markers, I will demonstrate in this section that the operator movement analysis for adverbial clauses proposed in section 3 can also apply to relative clauses. In relative clauses, the *-o* final relative marker is the relative operator base generated in the argument position for the relativized NP, most of the time with the same  $\Phi$ -feature as the relativized NP (except for the cases like (17a)). It is then raised to Spec,CP by C and post-syntactically cliticizes onto the highest head with more than one syllables.

In the *amba* construction, after the operator is raised to Spec,CP, it cliticizes onto C (*amba*) which is the highest head with more than one syllables.

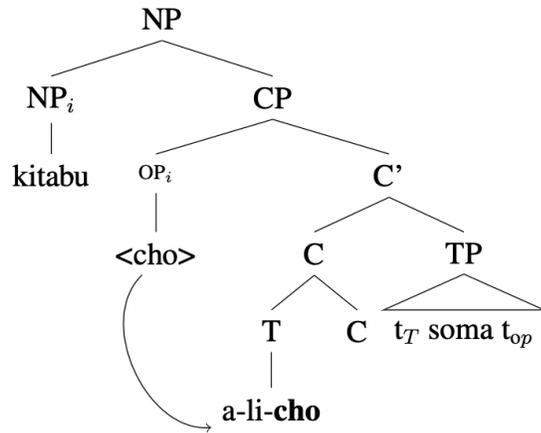
- (20) a. ki-tabu [CP amba-**cho** a-li-som-a]  
 7-book [CP amba-**7REL** 1SM-PST-read-IND]  
 the book that he read

- b. Proposed syntax of *amba*-relatives (final)



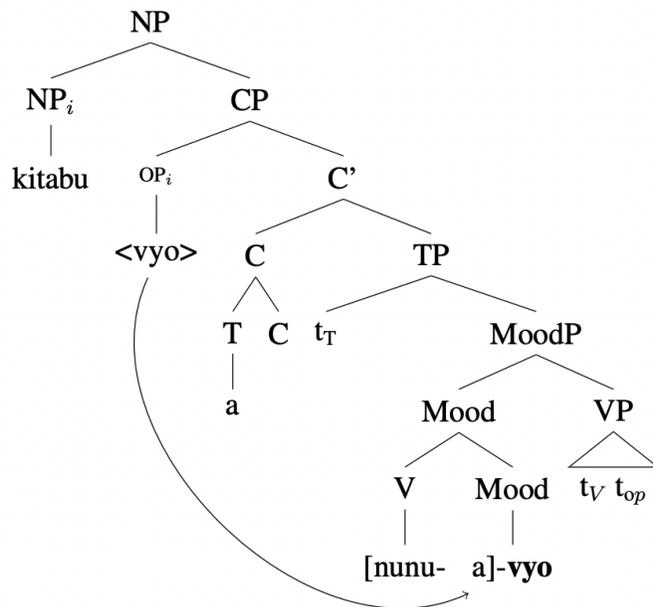
In the tensed *amba*-less construction, following Ngonyani (2006)'s analysis, T moves to C (Later I will show that T doesn't move to C in adverbial clauses), forming a complex C (T-C). In this case the highest head with more than one syllables is also C (*a-li*), onto which the raised operator cliticizes.

- (21) a. ki-tabu a-li-**cho**-som-a  
 7-book 1SM-PST-7REL-read-IND  
 the book that he read
- b. Proposed syntax of tensed relative (to be revised in section 5)



Finally, in the tenseless *amba*-less construction, I propose that T also moves to C, and the verb stem stays in situ, contrary to Ngonyani (2006)'s analysis that the verb stem moves to C. In this case, T is monosyllabic because it only bears the subject marker, thus after T moves to C, C is also monosyllabic. Then the highest head with more than one syllables is Mood. Thus, the raised operator cliticizes onto Mood.

- (22) a. vi-tabu a-nunu-a-vyo  
 8-book 1SM-buy-IND-8REL  
 The books he buys
- b. Proposed syntax of tenseless relatives (to be revised in section 5)



Note that derivations in (21b) and (22b) don't account for the cases where an overt subject is present in the relative clause. Since overt subject situates in Spec,TP, and T has moved to C, the overt subject will intervene between the T-C complex and the verb stem, resulting in the following ungrammatical morpheme order.

(23) Ungrammatical morpheme order when overt subject is present

- a. \*ki-tabu a-li-cho           Juma som-a  
7-book 1SM-PST-7REL Juma read-IND
- b. \*vi-tabu a    Juma nunu-a-vyo  
8-book 1SM Juma buy-IND-8REL

This problem is related to the treatment of post-verbal subjects in relative clauses and the word order asymmetry between relative clauses and adverbial clauses. I return to this issue in section 5, where I revise the analysis slightly to derive the correct position of overt subjects in relative clauses.

#### 4.4 Accounting for additional facts

In this section, I demonstrate that the analysis proposed here derives two additional facts about relative clauses, discussed in Ngonyani 2006. The first is that in tenseless negative relative clauses, instead of attaching to the end of the verb, the relative operator actually attaches to the end of the negative marker, before the verb stem.

- (24) a. \*ki-tabu tu-si-som-a-**cho**  
7-book we-NEG-read-IND-**7REL**
- b. ki-tabu tu-si-**cho**-som-a  
7-book we-NEG-**7REL**-read-IND  
the book which we do not read.  
Ngonyani (2006)

Since there is no tense marker in (24b), the position of the relative marker seemingly violates the generalization that the operator attaches to the end of the verb in tenseless relative clauses. My proposal states that the position of the operator is actually determined by the phonological property of the target of cliticization. Thus, if we posit that *tu-si* is a single disyllabic head for NEG, we can explain why the operator cliticizes onto this position even though the clause is tenseless.

Another fact is that some verbs only allow the *amba-* construction of relative clauses. An example is the verb "to have" in the present tense given in Ngonyani (2006).

- (25) a. m-toto a-na ki-tabu  
 3-child 1SM-with 7-book  
 the child has the book
- b. \*ki-tabu a-na-cho m-toto  
 7-book 1SM-with-7REL 3-child
- c. ki-tabu amba-cho m-toto a-na-cho  
 7-book amba-7REL 3-child 1SM-with-7OM  
 the book that the child has  
 Ngonyani (2006)

Notice that the present tense of this verb doesn't have a tense marker, thus T only bears a monosyllabic subject marker, and the verb stem is also monosyllabic. Notice that although *a-na* itself is disyllabic, it is formed by two separate monosyllabic heads, thus there is no available head in the CP domain for the operator to cliticize onto. Therefore, the *amba* construction is the only possible option.

## 5 Word order asymmetry between relative clauses and adverbial clauses

A difference between Swahili relative clauses and adverbial clauses is that in *amba*-less constructions, the subject must be post-verbal in relative clauses, while in adverbial clauses it can be pre-verbal. Relevant examples are repeated below.

- (26) ni-li-pig-a chafya [<sub>CP</sub> **Baraka** a-li-po-kat-a ki-tungu]  
 1sg-PST-hit-IND sneeze [<sub>CP</sub> **Baraka** 1SM-PST-SUB-cut-IND 7-onion]  
 I sneezed when Baraka cut the onion
- (27) a. vi-tabu [<sub>CP</sub> a-li-vyo-nunu-a **Juma**] ni ghali  
 8-book [<sub>CP</sub> 1SM-PST-8REL-buy-IND **Juma**] be expensive  
 The books Juma bought are expensive.
- b. \*vi-tabu [<sub>CP</sub> **Juma** a-li-vyo-nunu-a] ni ghali  
 8-book [<sub>CP</sub> **Juma** 1SM-PST-8REL-buy-IND] be expensive

While there are cases in which subject follows the verb in adverbial clauses, later in this section I show that this is not the same kind of phenomenon as the post-verbal subject in *amba*-less relative clauses.

- (28) ni-li-pig-a chafia [a-li-po-kat-a **Baraka** ki-tungu]  
 1SG-PST-hit-IND sneeze [1SM-PST-SUB-cut-IND **Baraka** 7-onion]  
 I sneezed when Baraka cut the onion

In this section, I adopt and extend Ngonyani (2006)'s MoodP topicalization analysis to account for the obligatory post-verbal subject in *amba*-less relative clauses, and argue that the word order asymmetry between these two clause types is caused by the truncated left periphery of adverbial clauses. Finally, I provide a separate account for post-verbal subjects in adverbial clauses, which is actually a general phenomenon in Swahili clauses.

### 5.1 MoodP topicalization in relative clauses

Recall that the partial operator movement analysis proposed in section 4.3 was not able to derive the correct morpheme order for relative clauses with overt subjects. Since T always moves to C in *amba*-less relative clauses, and the overt subject situates in Spec,TP, the overt subject intervenes between the T-C complex and the verb stem (29).

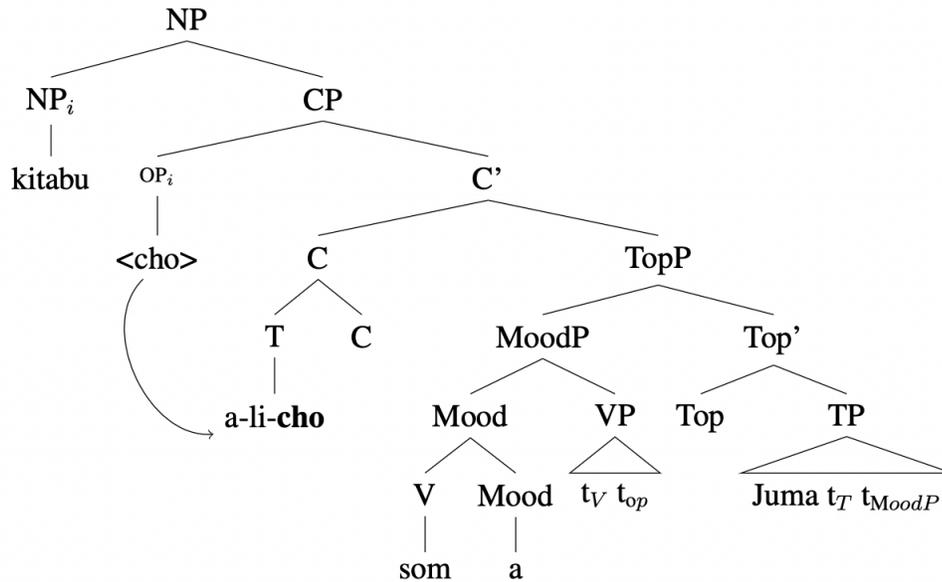
(29) Ungrammatical morpheme order when overt subject is present

- a. \*ki-tabu a-li-cho            Juma som-a  
   7-book 1SM-PST-7REL Juma read-IND
- b. \*vi-tabu a        Juma nunu-a-vyo  
   8-book 1SM Juma buy-IND-8REL

To address this issue, Ngonyani (2006) proposes MoodP topicalization, where the MoodP is moved to Spec,TopP, a position higher than TP and lower than C. Applying this movement to (21b), the correct morpheme order of (30a) is derived.

- (30) a. ki-tabu a-li-cho-som-a Juma  
 7-book 1SM-PST-7REL-read-IND Juma  
 the book that Juma read  
 Ngonyani (2006)

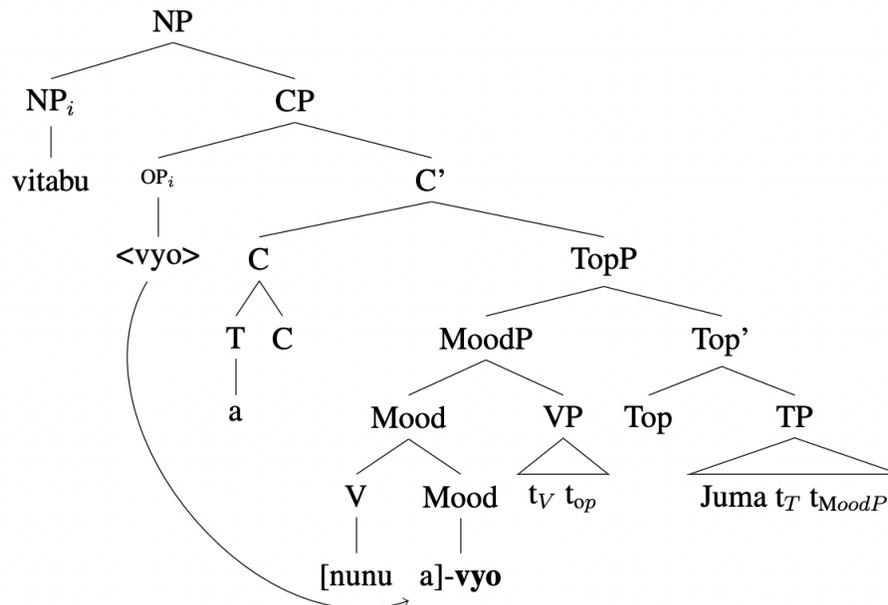
- b. Proposed syntax for tensed relatives (final)



Ngonyani (2006) only proposes this movement for **tensed** *amba*-less relative clauses, since in his analysis of **tenseless** *amba*-less relative clauses, the verb itself moves to C and consequently precedes the subject, rendering MoodP topicalization unnecessary. Recall, however, that under the account proposed here, tensed and tenseless relatives have a more uniform derivation — in particular, the both involve T to C movement, and neither involves movement V to T movement. (See (21b) and (22b)). For this reason, MoodP topicalization is necessary in both to derive the post-verbal subject position. The complete derivation of a tenseless relative under my account is given in (31b).

- (31) a. vi-tabu a-nunu-a-vyo Juma  
 8-book 1SM-buy-IND-8REL Juma  
 The books Juma buys  
 Ngonyani (2006)

- b. Proposed syntax of tenseless relatives (final)



At this point, I conclude that the obligatory post-verbal subject in *amba*-less relative clauses is the result of two obligatory movements in these constructions: T to C movement followed by MoodP topicalization. These two movements cause both T and Mood to precede the subject, while still adjacent to each other, manifested as the subject following the verb.

## 5.2 Obligatory clustering of T to C movement and MoodP topicalization

As shown previously, because T moves to C in *amba*-less relative clauses, MoodP topicalization is required to derive the correct morpheme order. In this section I argue that in Swahili clauses, T to C movement and MoodP topicalization are actually obligatorily clustered, meaning that it is impossible for one movement to take place while not the other.

It has been shown that in *amba*-less relative clauses, since T moves to C, MoodP must move to Spec,TopP in order to derive the correct morpheme order.

(32) T to C movement requires MoodP topicalization.

- a. \*kitungu a-li-cho Baraka [MoodP kat-a]  
 onion 1SM-PST-7REL Baraka [MoodP cut-IND]  
 (+T to C, -MoodP top.)
- b. kitungu a-li-cho-kat-a Baraka  
 onion 1SM-PST-7REL-cut-IND Baraka  
 The onion which Baraka cut  
 (+T to C, +MoodP top.)

On the other hand, in the *amba* construction of relative clauses, since C is already occupied, T doesn't move to C (20b). In this case, if MoodP moves to Spec,TopP, ungrammatical morpheme order will be derived.

(33) Absence of T to C movement precludes MoodP topicalization.

- a. \*ki-tabu amba-cho [TopP [MoodP som-a] [TP Juma [T a-li]]]  
 7-book amba-7REL [TopP [MoodP read-IND] [TP Juma [T 1SM-PST]]]  
 (-T to C, +MoodP top.)
- b. ki-tabu amba-cho Juma a-li-som-a  
 7-book amba-7REL Juma 1SM-PST-read-IND  
 the book Juma read.  
 (-T to C, -MoodP top.)

Thus, I propose the following generalization for Swahili clauses.

(34) *T to C movement and MoodP to Spec,TopP movement mutually entail each other in Swahili clauses: The presence of one requires the other and the absence of one precludes the other.*

In the next section, I show that this generalization is responsible for the word order asymmetry between relative clauses and adverbial clauses.

### 5.3 Accounting for the word order asymmetry

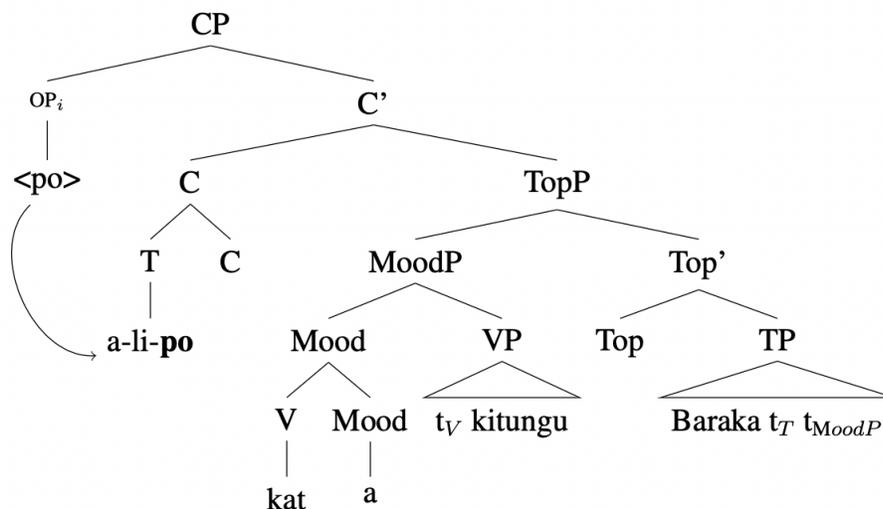
In this section, I show that the word order asymmetry is caused by the truncated left periphery of adverbial clauses. It has been noticed for English that adverbial clauses resist argument fronting.

- (35) a. \*When her regular column she began to write again, I thought she would be OK.
- b. When she began to write her regular column again, I thought she would be OK.

Haegeman (2010) accounts for this fact by proposing that the left periphery of adverbial clauses is truncated. It states that several functional projections, including TopP, don't exist in the left

periphery of adverbial clauses. I argue that this is true for Swahili adverbial clauses as well: that there is no Topic project in Swahili adverbial clauses. For this reason, MoodP topicalization is impossible in adverbial clauses. Since MoodP movement is absent, T doesn't move to C because of (34). Thus, T and MoodP have to stay in situ in adverbial clauses. The following examples demonstrate this fact by showing that if MoodP is topicalized and T moves to C in adverbial clauses, the sentence will be ungrammatical.

- (36) a. \**ni-li-pig-a*      *chafya* [<sub>CP</sub> *a-li-po-kat-a*                      *ki-tungu Baraka*]  
 1SG-PST-hit-IND sneeze [<sub>CP</sub> 1SM-PST-SUB-cut-IND 7-onion **Baraka**]  
 Intended: I sneezed when Baraka cut the onion.
- b. When MoodP is topicalized and T moves to C in adverbial clauses, the sentence will be ungrammatical.



Notice that when an object is present in the subordinate clause, MoodP topicalization will result in the order such that the verb precedes the object, and the object precedes the subject (VOS). (36) shows that this order is not allowed in adverbial clauses. However, Ngonyani (2006) shows that this order is allowed in relative clauses.

- (37) *wa-toto* [<sub>CP</sub> *a-li-o-wa-nunul-i-a*                      *vi-tabu Juma*] *wa-me-furahi*  
 2-child [<sub>CP</sub> 1SM-PST-2REL-2om-buy-APPL-IND 8-book Juma] 2SM-PERF-happy  
 The children for whom Juma bought books are happy  
 (Ngonyani, 2006)

In the relative clause in (37) in which the indirect object *watoto* is relativized, the direct object *vitabu* stays inside the relative clause, and it is moved with MoodP to a higher position than TP, preceding the subject *Juma*, resulting in the VOS order. Similarly, when the ungrammatical adverbial clause with the VOS order in (36a) acts as a relative clause, it becomes acceptable.

- (38) si-ku-u-*pend-a* u-le m-da [CP a-li-po-kat-a  
 NEG.1SG-NEG.PST-3OM-like-IND 3-DEM 3-time [CP 1SM-PST-16REL-cut-IND  
 ki-tungu Baraka]  
 7-onion Baraka  
 I didn't like the time when Baraka cut the onion.

Thus, (37) and (38) serve as additional evidence for the MoodP topicalization and T to C movement analysis for the post-verbal subject in relative clauses, and the contrast between (36a) and (38) is evidence that TopP is absent in Swahili adverbial clauses, which results in the word order asymmetry between adverbial clauses and relative clauses.

#### 5.4 Another kind of post-verbal subject phenomenon

As mentioned earlier, post-verbal subjects are also found in adverbial clauses, with example repeated below.

- (39) ni-li-pig-a chafia a-li-po-kat-a **Baraka** ki-tungu  
 1SM-PST-hit-IND sneeze 1SM-PST-SUB-cut-IND **Baraka** 7-onion  
 I sneezed when Baraka cut the onion

However, this is a fundamentally different phenomenon than the post-verbal subject in *amba*-less relative clauses which has been discussed. Post-verbal subject in *amba*-less relative clauses is analyzed with MoodP topicalization clustered with T to C movement, but this analysis cannot derive the word order in (39). As shown previously, this analysis would result in a VOS order whenever object is present in the subordinate CP, but (39) has the VSO order.

In fact, the VSO order is not only found in Swahili adverbial clauses. It is allowed in other kinds of Swahili clauses as well. The example below shows that matrix clauses also allow the VSO order.

- (40) a. Baraka a-li-kat-a ki-tungu  
 Baraka 1SM-PST-cut-IND 7-onion  
 Baraka cut the onion
- b. a-li-kat-a Baraka ki-tungu  
 1SM-PST-cut-IND Baraka 7-onion  
 Baraka cut the onion

On the other hand, Ngonyani (2006) pointed out an unexplained fact that in *amba* relative clauses, post-verbal subject is also allowed. This kind of post-verbal subject is also not the same kind as in *amba*-less relatives, because the analysis for the previous kind cannot apply here: T doesn't move to C in this case because C is occupied with *amba*. Thus, I posit that the relative clause in this case also has the VSO order, with the object relativized.

- (41) a. ki-ti amba-cho Juma a-li-nunu-a  
 7-chair amba-7REL Juma 1SM-PST-buy-IND  
 the chair that Juma bought
- b. ki-ti amba-cho a-li-nunu-a Juma  
 7-chair amba-7REL 1SM-PST-buy-IND Juma  
 the chair that Juma bought

I propose that the VSO order in these cases stems from the optionality of the EPP feature on T, disagreeing with an existing proposal that  $\Phi$ -features in Bantu must be associated with an EPP feature (Carstens, 2005). VSO order is derived whenever EPP is absent on T and the subject stays in situ.

Since the VSO order is available in all clause types, including relative clauses, it does not constitute an asymmetry between different clause types. Thus, I conclude that the word order asymmetry between adverbial clauses and relative clauses is solely caused by the truncated left periphery of adverbial clauses.

## 6 Conclusion

In this study, I have shown that Swahili adverbial clauses are very similar to relative clauses, a connection that is also observed in other languages. They are similar in the following ways: i) three characteristic constructions of relative clauses are all available in adverbial clauses. ii) the subordinator and the relative marker share the same form and iii) they can both modify an overt NP. This level of similarity motivates a unified analysis of these two clause types. I proposed that the subordinator and the relative marker are the same in nature, that they are operators base-generated in the position of the relativized constituent. The operator is raised to Spec,CP and cliticizes onto the highest head with more than one syllables, resulting in its three different surface positions. Notably, the operator is base-generated with its own  $\Phi$ -feature, and in relative clauses, this  $\Phi$ -feature isn't always the same as the  $\Phi$ -feature of the relativized NP. This mismatch is a key piece of evidence that challenges previous analyses of the relative marker as an agreeing complementizer, as well as the head raising analysis of Swahili relative clauses proposed by Ngonyani (2006).

I also account for the word order asymmetry between relative clauses and adverbial clauses. Specifically, I adopt and extend Ngonyani 2006's proposal that the obligatory VS and VOS word order in *amba*-less relative clauses result from MoodP topicalization clustered with T to C movement. I also show that the word order asymmetry is caused by the fact that adverbial clauses have a truncated left periphery (Haegeman, 2012). Since the Topic projection does not exist in the truncated left periphery of adverbial clauses, MoodP topicalization and T to C movements are impossible, resulting in the preference of SVO order in adverbial clauses.

While the proposed analysis in this work captures the similarity between Swahili relative clauses and adverbial clauses, as well as many other important facts, it fails to account for the ones presented in Ngonyani (2001) which support a raised head of the relative clause. Thus, a natural next step in this research is to attempt to understand the conflicting evidence for the derivation of

relative clauses in this language, and in particular, to test the hypothesis that both derivations are available.

## Acknowledgement

This work was advised by Dr. Asia Pietraszko, and it marks the start of my exciting journey of becoming a Linguist. I would like to thank Asia for her instructions, guidance, feedback and encouragements, especially for introducing me to the world of linguistic puzzles. Being her student was the luckiest thing happened to me during my difficult time. I would also like to thank Drusilla Talawa, my language consultant, for sharing her beautiful language and patiently answering my endless questions. Additionally, I'd like to thank our imaginary friend Baraka, who has cut excessive amount of onions in this study.

## References

- Ashton, Ethel O. 1947. *Swahili grammar, including intonation*. Longmans.
- Baker, Mark. 1985. The mirror principle and morphosyntactic explanation. *Linguistic inquiry* 16:373–415.
- Barrett-Keach, Camillia N. 1980. The syntax and interpretation of the relative clause construction in Swahili. Doctoral Dissertation, University of Massachusetts Amherst.
- Barrett-Keach, Camillia N. 1986. Word-internal evidence from Swahili for Aux/INFL. *Linguistic Inquiry* 17:559–564.
- Bhatt, Rajesh, and Roumyana Pancheva. 2017. Conditionals. In *The Wiley Blackwell companion to syntax*, ed. Martin Everaert and Henk C. Van Riemsdijk, 1–48. John Wiley & Sons, Inc. Hoboken, NJ, USA, 2 edition.
- Bianchi, Valentina. 2011. *Consequences of antisymmetry*. De Gruyter Mouton.
- Carstens, Vicki. 2005. Agree and EPP in Bantu. *Natural Language & Linguistic Theory* 23:219–279.
- Chomsky, Noam. 1995. *The minimalist program*. MIT Press.
- Declerck, Renaat HC. 2003. *When-clauses and temporal structure*. Routledge.
- Demirdache, Hamida, and Myriam Uribe-Etxebarria. 2004. The syntax of time adverbs. In *The syntax of time*, ed. Jacqueline Guéron and Jacqueline Lecarme, 143–180. The MIT Press.
- Ermisch, Sonja. 2012. Relative clause formation in the African context: The case of Swahili. *Gastvortrag an der Loránd-Eötvös-Universität Budapest* 10:2012.
- Geis, Michael Lorenz. 1970. Adverbial subordinate clauses in English. Doctoral Dissertation, Massachusetts Institute of Technology.
- Gregersen, Edgar. 1967. *Prefix and pronoun in Bantu*. Waverly Press.
- Haegeman, Liliane. 2010. The internal syntax of adverbial clauses. *Lingua* 120:628–648.
- Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena, and composition of the left periphery*, Vol. 8 of *The Cartography of Syntactic Structures*. Oxford University Press.
- Henderson, Brent Mykel. 2004. PF evidence for phases and distributed morphology. In *Proceedings of NELS*, Vol. 34, 255–266.
- Henderson, Brent Mykel. 2006. The syntax and typology of Bantu relative clauses. Doctoral Dissertation, University of Illinois at Urbana-Champaign.
- Hetterle, Katja. 2015. *Adverbial clauses in cross-linguistic perspective*. De Gruyter Mouton.

- Kayne, Richard S. 1994. *The antisymmetry of syntax*. MIT press.
- Kinyalolo, Kasangati KW. 2003. Limiting the scope of V-movement in Kilega. Handout of a talk given at Stony Brook, New York.
- Lindfors, Anna-Lena. 2004. Tense and aspect in Swahili. Ms., Uppsala University.
- Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic inquiry* 37:69–109.
- Mpiranya, Fidèle. 2014. *Swahili grammar and workbook*. Routledge.
- Ngonyani, Deo. 2001. Evidence for head raising in Kiswahili relative clauses. *Studies in African Linguistics* 30:60–73.
- Ngonyani, Deo. 2006. Attract F and verbal morphology in Kiswahili. *The Linguistic Review* 23:37–68.
- Schadeberg, Thilo C. 1989. The three relative constructions in Swahili (Kisanifu). In *Le swahili et ses limites: ambiguïté des notions reçues*, ed. Marie-Françoise Rombi, 33–40. Editions Recherche sur les Civilisations.
- Thompson, Sandra A, Robert E Longacre, and Shin Ja J Hwang. 1985. Adverbial clauses. In *Complex constructions*, ed. Timothy Shopen, Vol. 2 of *Language typology and syntactic description*, 171–234. 2 edition.
- Zwart, Jan-Wouter. 1997. A short note on relative constructions in Lega and Swahili. Ms., University of Groningen.

# THE DIPHTHONGIZATION OBSERVATION: AN ANALYSIS OF THE “INDIE VOICE” PERSONA

SARAH SMITH  
*University of Rochester*

## 1 Introduction

This paper aims to do the following: 1) describe the phenomenon of “indie voice” through a linguistic lens, touching on its public perception; 2) discuss previous sociolinguistic studies focused on style within the frame of singing and how those have inspired the current paper; 3) analyze data on indie-pop singer Halsey’s diphthongization of certain vowels when singing, and compare these diphthongs to the “pure” vowels she uses in normal speech; and 4) argue that this diphthongization—and by extrapolation, many other indie voice sound changes—is indexical of a specific “indie persona” by calling on and adapting performance- and persona-based theories of style-shifting.

## 2 Background

### 2.1 Sociolinguistics and Song

Spoken language and sung music are not isolated forms of language use. There are many overlapping factors that connect the variation observed in speech to that of singing, the most relevant for the purposes of this paper being register. Schilling-Estes defines a register to be a “readily identifiable speech variety that individuals use in specific, well-defined speech situations” (Schilling-Estes, 1998). Registers are subsumed within the realm of style, as they can also be viewed as a set of stylized variables bunched together to form an identifiable variety of speech. Registers also exist in music, and are typically genre-specific, as illustrated by Karen from Ace Linguist’s explanation that “you use very different vowels for classical operatic singing than you do when singing a bluesy rock song” (Ace Linguist, 2021). Singers commonly utilize a different register when singing from their typical spoken vernacular, sometimes picking

up forms in their sung accent that are not present at all in their everyday speech. This is not always a conscious decision, rather is sometimes physiologically necessary due to the mode of vocalization that a particular song requires. Articulations may be nudged around in the vocal tract away from their standards in order to produce a sound with more resonance, or one that matches the aural qualities of the instrumental better (Ugwu, 2015). This is not to say that sound changes in singing are strictly coming from physiological demand. Just as style-shifting in speech mostly happens due to social motivations, this can also be the case for song.

Spoken and sung language are not equivalent, however. Sung speech has an affordance to creativity that spoken speech cannot accommodate in the same way. Karen from *Ace Linguist* claims that “sung speech gives us an opportunity to witness experimentation and playfulness in the low-stakes world of music and art” (*Ace Linguist*, 2021). Music both sets up constraints, such as the need to follow a melody and rhythm, and supports experimentation due to its malleability, which are facets that come together to allow artists to come up with novel productions unheard in speech that work to not just suit the song but enhance it. These novel productions can systematize to create new linguistic repertoires, which over time and across singers may come to denote a particular genre or identity.

The connections between spoken and sung language, along with music’s ability to create novel linguistic identities, means that there have been multiple studies focused on sung speech. There have, of course, been considerations of popular song language from a musicological perspective, such as Blake’s description of timbre and Burkhart’s discussion of paralinguistic musical features, both explaining how these musical aspects index aesthetic values (Blake, 2012; Burkhart, 2017). But Trudgill’s seminal paper on the localized accents of British punk-rock singers, which examined their use of a low-prestige southern English accent in comparison to British pop singers’ “Americanized” features (Trudgill, 1997), served as the catalyst for many further sociolinguists curious about sung dialect. Simpson expanded upon Trudgill’s findings, defining more clearly the “American” features that British pop singers imported and how those differed from the singers’ spoken styles, as well as explaining that the stereotypically “American” features “do not always tally with any specific variety of American English” (Simpson, 1999). Beal connected this work to the concepts of performance and persona (to be discussed in Section 4) through doing a similar examination of the Arctic Monkeys’ use of local features (Beal, 2009). Jansen and Westphal took a different approach and looked at pop singer Rihanna’s use of Caribbean English Creole features in her music and how that highlighted her identity (Jansen & Westphal, 2017). Lastly, Jones et al. hypothesized on the articulatory placement of indie voice, which most closely relates to this study (Jones et al., 2017).

## 2.2 A Game of Telephone

As described in Trudgill (1997), Simpson (1999), and Beal (2009), Americans have dominated the popular music industry for decades; and in an attempt to replicate and associate themselves with the success of the American market, many artists in other English-speaking countries copied the most salient linguistic features of American accents and used those features in their own work. As the decades went on, with more and more mainstream singers all utilizing the same accent, those features began to index being “mainstream” rather than being “American”. This is when, in the 70’s, British punk-rock bands began to embrace their local Southern English accent and use that in their music instead. Similarly, Australian hip-hop artists tended to stick to their local voice rather than accommodating to the American pronunciation model. In both cases, the

rationale seems to be that using one's own vernacular, rather than changing one's voice to fit the same accent everyone in the corporate mainstream is using, asserts a genuine, "keeping it real" attitude. Jansen and Westphal describe this "going local" as "an active choice against an American-influenced mainstream and towards linguistic and creative independence and authenticity" (Jansen & Westphal, 2017). As these "down to earth" musicians became popular internationally, other vernaculars were introduced into the popular music sphere, and in particular many low-prestige Australian and British features were picked up and imitated by other counterculture artists. Then they too were copied, features found across multiple accents were converged, and these non-American linguistic features were spread across and became marked aspects of various alternative genres. This game of counterculture musical accent telephone may be (part of) how the phenomenon of "indie voice" came about (Ace Linguist, 2021).

### 2.3 Singing in Cursive

Indie music as a genre is characterized by an anti-mainstream ideology and an association with smaller record labels rather than huge management systems. It is reasonable to consider the indie movement a descendent of the punk values that sparked the telephone game described in the previous section, and therefore is also fair to assume that the accent used in indie music would also follow from the linguistic features derived from those values. "Indie Voice," also referred to as "singing in cursive" or "hip singing," is a particular register of sung speech used by many musicians in the indie genre. This style attracted large-scale attention in 2015, when a BuzzFeed article was written about singer Selena Gomez's use of it in her song "Good for You" (Ugwu, 2015). Not only was this style noticed, but it was also hugely criticized and mocked for sounding excessively ornate and being difficult to understand (Ace Linguist, 2021). The overall goal of indie voice is to share a feeling of intimacy, individuality, and vulnerability. Various linguistic features are characteristic of indie voice, such as a quiet breathy tone, lack of aspiration, extensive vocal fry, and shakiness. There are also distinct vowel choices which are particularly salient to listeners. Certain commonly remarked upon simple vowel changes are as follows:

/aɪ/ → [ɑɪ]                      /ʌ/ → [a]                      /eɪ/ → [æɪ]                      /æʊ/ → [aʊ]

But what's most unique to the register, as well as most ridiculed by listeners, is its distinctive substitution of front-rising diphthongs in the place of most any standard monophthong, examples being:

/ɛ/ → [ɛɪ]                      /ʌ/ → [ʌɪ]                      /ɔ/ → [ɔɪ]

In her general guide to the observed sounds of indie voice, Karen from the blog Ace Linguist reports that, "Diphthongization can be very obvious or extremely subtle, but either way the vowels no longer sound pure. These diphthongs are closing diphthongs – they go from a low vowel to a high vowel. The one exception is /ʊ/ → [ʊɪ], where the tongue stays at the same height as it moves forward" (Ace Linguist, 2021). Nearly all diphthongized vowels move specifically towards an /i/ sound, and most occur before a coronal consonant, which lead Jones et al. to believe that "the diphthongs are prolonged audible transitions between the tongue's vocalic position and the articulatory target for the following consonant" (Jones et al., 2017).

This characteristic diphthongization will be the focus for the data analysis portion of this paper. This is for two main reasons. First, it is perceptually salient and easy to pick out. Second, it is a feature that is not already indexical of some other genre, identity, or location. Many of the

individual vowel changes come from British and Australian dialects, and the overall soft voice quality features are taken from R&B and folk music, but diphthongization is purely an indie voice characteristic.

## 3 Data Analysis

### 3.1 The Variable and The Question

What we're concerned with is the addition of a short, secondary vowel to a standardly monophthong vowel, henceforth called diphthongization. The variable is the diphthongized quality of a vowel, and the variants are whether that quality is present or not. The envelope of variation encompasses all vowels except for the high-front /i/ and /ɪ/, given those can't slide to "close off" any higher or more forward. Measurement of this variable was done both impressionistically (whether or not a diphthong is present) and quantitatively (using formant values). A visual comparison of a diphthongized /ʌ/ and a monophthong /ʌ/ is in Figure 1.

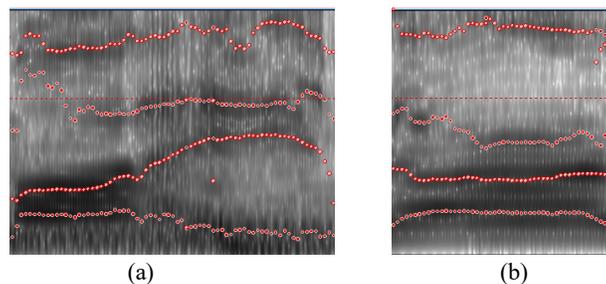


Fig 1. Spectrograms for (a) the diphthongized [ʌ] in “much” and (b) monophthong [ʌ] in “clubs”. The formant shift to the [ɪ] is clearly present in (a) and not in (b).

The question: Will American indie(-pop) singer Halsey, who is commonly quoted as being a user of indie voice, produce this extensive diphthongization in her singing? Will it be in the manner and frequency described (closing diphthongs, all the time)? Will this differ from how she pronounces vowels in everyday speech? I hypothesize that Halsey will indeed often use extensive diphthongization by adding a short /ɪ/ to multiple vowels that are pronounced as monophthongs in her spoken language.

### 3.2 Method

Two clips were obtained from Australian radio station Triple J's “Like A Version,” which each include audio of Halsey both speaking and singing. These clips were both taken on the same day, with the same audience (two interviewers and two musicians) present. The total length of both clips together comes to 10 minutes 21 seconds, with around 2 minutes of that time devoted to talking and 8 minutes spent singing. The speaking and singing portions were extracted from each other, put into Praat, and hand-transcribed and annotated on the word level. Any audio sections where voices overlapped or there were purely instrumentals with no singing were deleted. The annotated files were put through DARLA's Semi-Automated Alignment and Extraction suite to conduct mean formant analyses to allow visual description of the changing vowels (Fig. 2). DARLA does not automatically provide the second formant values for diphthongs, so every

diphthong’s two sets of formants were calculated using Praat one at a time. Files were then auditorily examined to acquire token counts of each type of vowel, with a focus on catching any uses of *i*-diphthongization.

### 3.3 Results and Analysis

Token counts for each standard type of vowel, as well as how it was actually pronounced in each speech setting, are shown in the chart below. In this and all future analysis, “standardly monophthongs” refers to the vowels {*ɑ*, *æ*, *ʌ*, *ɔ*, *ə*, *ɜ*, *ɛ*, *ɪ*, *i*, *ʊ*, *u*} and “standardly diphthongs” refers to {*aʊ*, *aɪ*, *eɪ*, *oʊ*}.

	Speaking		Singing	
	Pronounced as Monophthong	Pronounced as Diphthong	Pronounced as Monophthong	Pronounced as Diphthong
Standardly Monophthongs	337	0	654	38
Standardly Diphthongs	24	69	102	179

The boxes of interest in the charts are those of a standard monophthong pronounced as a diphthong (diphthongization), for which Halsey had no tokens when speaking, but 38 when singing. 38 is not a particularly large number, but it is far more than zero. To confirm, a two-sample t-test assuming equal variances was run on the standardly monophthong data to determine whether the diphthongization difference between speech settings was significant. Means were acquired by coding all instances of diphthongization as 1 and standard monophthong pronunciations as 0. The resulting p-value was  $p = 0.0000108$ , which is much less than  $\alpha = 0.05$ , so we can conclude that the difference is significant, and it is her indie voice singing environment that is inspiring diphthongization.

Not every vowel underwent diphthongization. Only /*ɑ*/, /*æ*/, /*ʌ*/, /*ɔ*/, /*ə*/, and /*ɛ*/ had diphthongized tokens, meaning that /*ɜ*/, /*ʊ*/, and /*u*/ were excluded. This makes sense, as /*ɜ*/ is sometimes considered to be a diphthong in the first place, so it shouldn’t undergo extra diphthongization. Both *ʊ* and *u* are already high vowels. They do not have the ability to be raised further, only fronted. As mentioned in Section 2.3, diphthongization should be possible on /*ʊ*/, so either it was just unfortunate that Halsey’s audio didn’t have any tokens of that, or diphthongization is only available for vowels that can be both fronted and raised, but exploring that theory further is outside the scope of this paper.

For a visual representation of the sound changes happening across different vowel conditions, mean formant values for all spoken vowels, all sung vowels, and all diphthongized vowels were put into the NORM Vowel Normalization and Plotting Suite to create the vowel space maps in Figure 2. NORM requires Arpabet vowel notation rather than IPA, so that is what is shown here. Conversions between the two systems are available online.

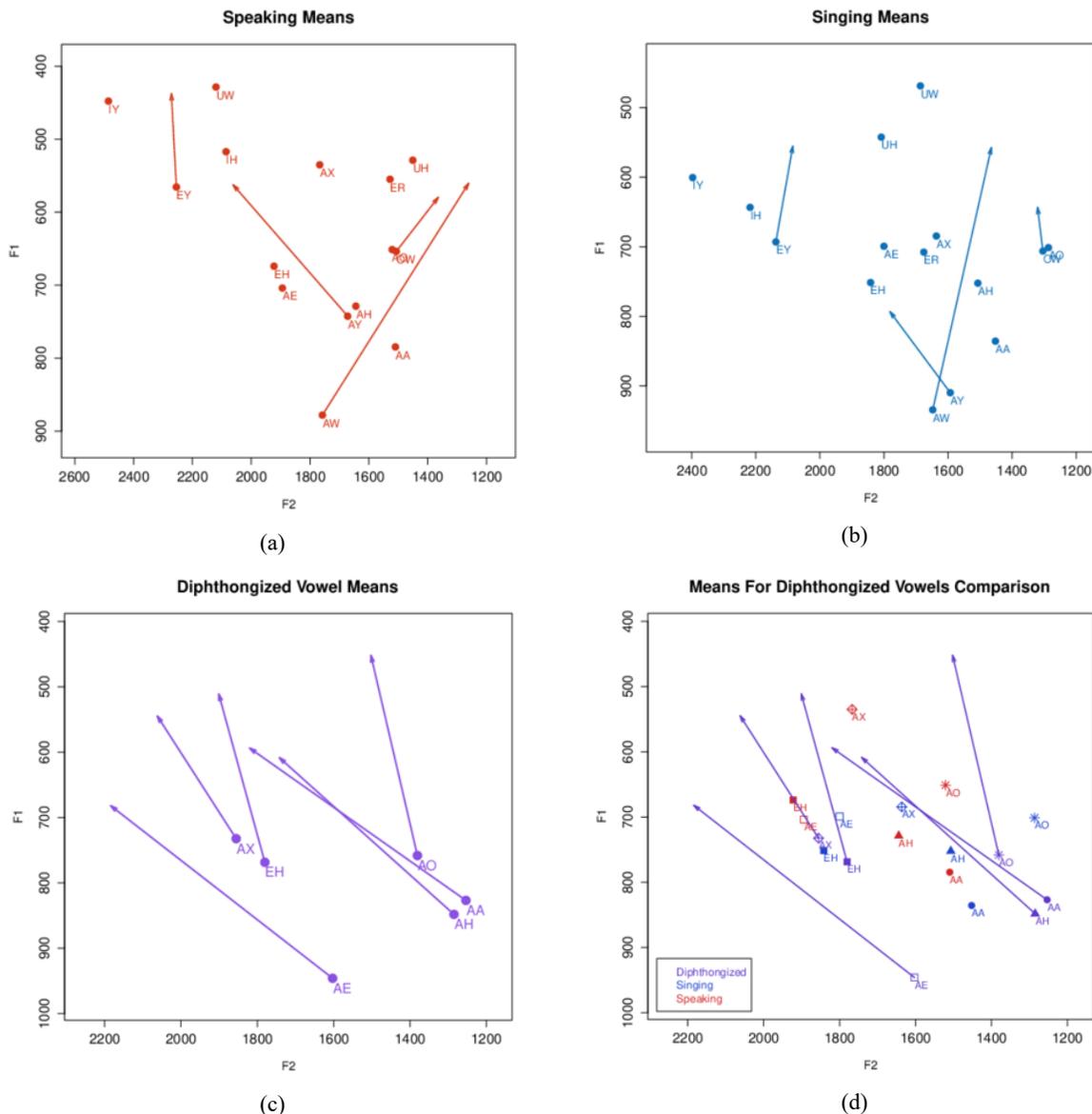


Fig 2. Vowel spaces for (a) all vowel means from Halsey's speaking audio, (b) all vowel means (excluding diphthongized ones) for her singing audio, (c) means for the diphthongized vowels from her singing audio, and (d) comparison between the placement of the diphthongized vowels and their standard monophthong counterparts from speech and song.

It is clear from these plots that vowels are generally being somewhat lowered and backed in song from their positions in speech. What's more interesting is how the diphthongized vowels are nearly all starting lower than both their spoken and sung counterparts. Their moving higher and to a more fronted position is a given considering the definition of i-diphthongization, but this initial lowering is an effect I did not foresee, and it would be interesting to explore the reasons behind it in further research. My tentative explanation is that this happens in order to make the glide from the initial vowel to /i/ more distinct and audible. Now that it has been established that diphthongization and other indie voice sound changes are indeed happening in the manner described in Section 2.3, it's important to discuss why they might be happening.

## 4 Connections to Theory

### 4.1 Previous Theories

Various scholars have proposed theories as to why style shifting happens in music. Beal considered the possibility of Giles and Smith's accommodation theory, which states that "speakers modify their accents to either converge to or diverge from those of their interlocutors" (Beal, 2009). The issue that arises here is that in the performance of a song, a singer has no interlocutors. It is not a conversation that is taking place. Bell's audience design theory attempts to remedy this problem by placing the audience in the position of interlocutors, so singers would be accommodating to the accent of their audience. This theory is also problematic, since artists tend to stick to a particular style across all of their music but perform to both many different populations (tours, concerts) and sometimes no audience (recording sessions). Thus, they cannot be accommodating towards or against both every possible audience and no audience in songs that are sung in the same style across performances.

Another possible theory is that of "appropriateness". Trudgill explains this concept by saying that "Different situations, different topics, different genres require different linguistic styles and registers. The singing of pop music in this way, it could be argued, is no different from vicars preaching in the register appropriate to Church of England sermons, or BBC newsreaders employing the variety appropriate for the reading of the news" (Trudgill, 1997). He then deprecates this theory, however, by claiming that it is not enough to simply say a certain register is appropriate for a genre, but it must be explained why the specific features used are considered appropriate.

There is also consideration of the articulatory settings used by singers of different genres. Jones et al. illustrates that "The articulatory settings associated with a given singing style are known to have noticeable, systemic effects on the acoustics of the singer's pronunciation" (Jones et al., 2017). This is the prospect that they investigated in their 2017 study, defending the hypothesis that indie voice is the result of a pharyngealized articulatory setting. Even they admit that the phenomenon is not just physiologically driven, and that "while indie-pop's distinctive diphthongs may originally have been a by-product of an articulatory setting, they have since been adopted as part of a musical style" (Jones et al., 2017), that is, the sound changes became an overt aspect of the way users pick up and identify the style, rather than just being a symptom of a particular articulation.

None of these theories have provided us with a complete picture of the reasoning behind the shift to indie voice. I propose that a mixture of Jones et al.'s articulation theory, various scholars' concept of persona-driven style, and Schilling-Estes' and Coupland's concept of performed language all work together to, respectively, target what Simpson says are the three interrelated factors that influence singing style, "the nature of the perceptual linguistic model aspired to, the nature of the pop and rock bands who adopt it and the nature of the bands' targeted audience" (Simpson, 1999).

### 4.2 Performance

In her paper on Ocracoke English, Schilling-Estes introduces the register of Performance Speech, which is "associated with speakers' attempting to display for others a certain language or

language variety, whether their own or that of another speech community” (Schilling-Estes, 1998). Performance speech targets specific features that are perceptible from the target dialect. She explains this by saying that “When speakers attempt to “put on” a dialect for an audience, they have available to them only those features they can perceive; further, there is evidence that the greater perceptual awareness speakers have of a given language feature (whether this awareness is at the conscious level or not), the greater the extent to which the feature will figure in their demonstrations and discussions of the language variety in question” (Schilling-Estes, 1998). This tendency to lean into particularly noticeable features lines up with the passing down/imitation chain of features described in Section 2.2, justifying why perceptible changes (like diphthongization) become characteristic among users of the indie voice register.

Singing comes under the umbrella of what Coupland calls “high performance”. It is planned (not impromptu like standard speech), intentionally stylized, and aimed for entertainment rather than conversation. Singers are performing the speech of a character (that may or may not be themselves), that affords a “crucial relational dynamic between a performer and the individuated audience member, who may align with or even feel transcendently drawn into the performer’s own identity” (Coupland, 2011). This relational dynamic inspires “‘higher order indexicalities’ – awareness that a certain stylistic variant operates as an index for a certain social meaning” (Bell & Gibson, 2011). Through these performer-audience relations, the stylized form of the performance becomes enregistered (“the process by which a style becomes engraved in the public mind as indexing certain social personas or ‘characterological figures’” (Bell & Gibson, 2011) as indexical of an indie singer. Indie artists are not converging with anyone, since there is no dialect in the spoken world that uses all of the features of indie voice. Rather, during performances, they are differentiating themselves from everyone, to try to push their special hippie image.

### 4.3 Persona and Indexicality

Le Page’s “Acts of Identity” theory, where linguistic behavior is motivated by attempts to “resemble as closely as possible those of the group or groups with which from time to time we [speakers] wish to identify” (Beal, 2009), seems more encompassing than those discussed in Section 4.1. In terms of this theory, indie singers are modifying their pronunciation when singing towards that of previous users of indie voice, a group with which they wish to identify. Beal claims that “Within the language-ideological framework, linguistic features are seen to become associated with social values so that they acquire indexical meanings. These meanings can change through time, and different groups in a community may attribute different values to the same linguistic feature” (Beal, 2009). What style shifting to indie voice appears to be doing is pushing forward a speaker’s “projected social role and persona” (Simpson, 1999) that, in this genre, aims to project a raw, uncensored image with a “principle of transparency in performance that suggests real and deep experience” and a social “premium on individual character over established convention” (Coupland, 2011; Ugwu, 2015).

How do indie singers achieve this image? In Zhang’s account of persona-driven language, he shows that “To create a distinctive style, individuals draw upon resources that are accessible to them. However, they do not pick up stylistic material in a free-wheeling fashion. The creation of style can be seen as a process of “bricolage” (Levi- Strauss 1966) in which the “bricoleur” selects from a limited and preexisting set of materials at hand and arranges them into a meaningful ensemble” (Zhang, 2005). Following this framework, indie voice is a bricolage of many sound

changes from no one clear source. The use of some of these features (vocal fry, breathy voice) is already culturally indexical of a personal, authentic tone. The indie singers taking those known aspects, as well as copying the noticeable vowel features of previous indie singers and other counterculture musicians and putting everything together allows them to portray a down-to-earth (as opposed to popular, celebrity), “individualistic” persona that associates them with the indie community. Although it may seem strange for such a hodgepodge bundle of features to imply the indie persona in particular, Bell and Gibson clarify that “Indexes do not directly resemble the referent but reference it through association and co-occurrence—smoke is an index of fire because the two co-occur. Linguistic variation comes to have social meaning through the co-occurrence of certain variants with perceived categories of speaker, or with certain genres of speech event” (Bell & Gibson, 2011). As demonstrated in the data above, the co-occurrence of diphthongization and other indie voice features with Halsey’s performance, as opposed to the lack of co-occurrence in her speaking, allows her singing to be indexical of the indie voice persona.

## 5 Conclusion

This paper has covered the history of sociolinguistics in popular song, the description and analysis of indie voice both generally and specifically looking at diphthongization, and sociolinguistic theories as to why indie voice is used. Goals for further research in formally describing indie voice may be determining whether measuring formants within a song, given the amount of intervening noise provided by instrumentals, is a reliable methodology in order to not rely on impressionistic classification, utilizing more data (longer speech, more songs), searching if there is correlation between indie voice and topic of communication, and looking at multiple users of indie voice and see what features they share with each other, and which they don’t.

## References

- Ace Linguist (2021, May 3). *Dialect Dissection: Indie Voice/Cursive Singing - The Definitive Post*. Ace Linguist; Blogger. <https://www.acelinguist.com/2021/05/dialect-dissection-indie-voicecursive.html>
- Beal, Joan (2009). “You’re Not from New York City, You’re from Rotherham.” *Journal of English Linguistics*, 37(3), 223–240. <https://doi.org/10.1177/0075424209340014>
- Bell, Allan and Andy Gibson (2011). Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics*, 15(5), 555–572. <https://doi.org/10.1111/j.1467-9841.2011.00517.x>
- Blake, David K. (2012). Timbre as Differentiation in Indie Music. *Music Theory Online*, 18(2). <https://mtosmt.org/issues/mto.12.18.2/mto.12.18.2.blake.php>
- Burkhart, Benjamin (2017) Popular Music Analysis and Social Semiotics: The Case of the Reggae Voice. In: Merrill J. (eds) *Popular Music Studies Today*. Systematische Musikwissenschaft. Springer VS, Wiesbaden. [https://doi.org/10.1007/978-3-658-17740-9\\_4](https://doi.org/10.1007/978-3-658-17740-9_4)
- Coupland, Nikolas (2011). Voice, place and genre in popular song performance. *Journal of Sociolinguistics*, 15(5), 573–602. <https://doi.org/10.1111/j.1467-9841.2011.00514.x>

- Jansen, Lisa and Michael Westphal (2017). Rihanna Works Her Multivocal Pop Persona: A Morpho-syntactic and Accent Analysis of Rihanna's Singing Style: Pop culture provides rich data that demonstrate the complex interplay of World Englishes. *English Today*, 33(2), 46-55. <http://dx.doi.org/10.1017/S0266078416000651>
- Jones, Colin, Schellenberg, Murray and Bryan Gick. (2017). Indie-Pop Voice: How a Pharyngeal/Retracted Articulatory Setting May Be Driving a New Singing Style. *Canadian Acoustics*, 45, 180-181.
- Schilling-Estes, Natalie (1998). Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society*, 27(1), 53–83.
- Simpson, Paul (1999). Language, Culture and Identity: with (another) Look at Accents in Pop and Rock Singing. *Multilingua*, 18(4), 343-367. <https://doi.org/10.1515/mult.1999.18.4.343>
- triple j (2016a). Halsey - "Hold Me Down" (live for Like A Version) [YouTube Video]. In *YouTube*. [https://www.youtube.com/watch?v=8R\\_drHL6o9E](https://www.youtube.com/watch?v=8R_drHL6o9E)
- triple j (2016b). Halsey covers Justin Bieber "Love Yourself" for Like A Version [YouTube Video]. In *YouTube*. [https://www.youtube.com/watch?v=\\_1pUqmaPsFA&t=16s](https://www.youtube.com/watch?v=_1pUqmaPsFA&t=16s)
- Trudgill, Peter (1997). Acts of Conflicting Identity: The Sociolinguistics of British Pop-song Pronunciation. In *Sociolinguistics* (pp. 251–265). Macmillan. <https://link.springer.com/book/10.1007%2F978-1-349-25582-5>
- Ugwu, Reggie (2015, October 6). *Selena Gomez's "Good For You" And The Rise Of "Indie Pop Voice."* BuzzFeed News; BuzzFeed News. <https://www.buzzfeednews.com/article/reggieugwu/what-is-indie-pop-voice>
- Zhang, Qing (2005). A Chinese yuppie in Beijing: phonological variation and the construction of a new professional identity. *Language in Society*, 34(3), 431-466.

# THE LONG-DISTANCE ‘TAZIJI’ REVISITED: EXPERIMENTAL EVIDENCE FOR INHERENT NON-LOCAL BINDINGS\*

YUHANG XU

JEFFREY T. RUNNER

*University of Rochester*

## 1 Introduction

Chinese reflexives provide a unique typological test bed for evaluating theories of anaphora in theoretical linguistics and psycholinguistics (see recent reviews, e.g., Jäger, Engelmann, and Vasishth 2015; Charnavel et al. 2017). It is well documented that there are at least two types of reflexives in Mandarin Chinese (e.g., Y.-H. Huang 1984; L. Xu 1993; Pan 1998; Yu 2000; C.-T. J. Huang, Li, and Li 2009; Hu 2019), the compound/polymorphemic reflexive *taziji* (pronoun + “self”, similar to *himself* or *herself* in English), and the bare/monomorphemic reflexive *ziji* (“self”). The pronoun *ta* and the compound reflexive *taziji* are often considered to be similar to their English counterparts *him/her* and *himself/herself* which strictly follow the binding principles (e.g., Chomsky 1981) while the bare reflexive *ziji* is thought to belong to the long-distance reflexive family which are “exempt” from Principle A and are subject to another set of constraints (e.g., Tang 1989; Pan 1998).

We present in (1) the original examples<sup>1</sup> and reported judgments from a widely used textbook, *The syntax of Chinese* (C.-T. J. Huang, Li, and Li 2009), to illustrate the syntactic behavior of Chinese anaphors as generally understood by the field. According to C.-T. J. Huang, Li, and Li (2009), the pronoun *ta* as in (1a) can only take, abiding by binding principles, the non-local subject *Zhangsan* as its antecedent. Whereas the reflexive *taziji* as in (1b), must only take the local subject *Lisi* as its antecedent. On the other hand, for the bare reflexive *ziji*, both local and long-distance binding options are available as shown in (1c).

---

\*We would like to thank Prof. Ash Asudeh for his insightful comments and suggestions that improved this paper.

<sup>1</sup>The examples are presented in Latin transcriptions in the textbook. The Chinese scripts are based on a translation version of the same textbook (C.-T. J. Huang, Li, and Li 2013).

- (1) C.-T.J.Huang et al. (2009, pp. 330,333).
- a. 张三 知道 李四 老 批评 他  
 Zhangsan<sub>i</sub> zhidao Lisi<sub>j</sub> lao piping ta<sub>i/\*j</sub>  
 Zhangsan<sub>i</sub> know Lisi<sub>j</sub> incessantly criticize ta<sub>i/\*j</sub>  
 ‘Zhangsan knows that Lisi criticizes him all the time.’
- b. 张三 知道 李四 老 批评 他自己  
 Zhangsan<sub>i</sub> zhidao Lisi<sub>j</sub> lao piping taziji<sub>\*i/j</sub>  
 Zhangsan<sub>i</sub> know Lisi<sub>j</sub> incessantly criticize taziji<sub>\*i/j</sub>  
 ‘Zhangsan knows that Lisi criticizes himself all the time.’
- c. 张三 知道 李四 老 批评 自己  
 Zhangsan<sub>i</sub> zhidao Lisi<sub>j</sub> lao piping ziji<sub>i/j</sub>  
 Zhangsan<sub>i</sub> know Lisi<sub>j</sub> incessantly criticize ziji<sub>i/j</sub>  
 ‘Zhangsan knows that Lisi criticizes self all the time.’

Also similar to English, Mandarin Chinese has different forms for the masculine and feminine third person singular pronouns and reflexives. However, they are homophones that are distinguishable only in their orthography<sup>2</sup>. For example, besides the gender neutral reflexive *ziji* 自己, the masculine and feminine pronoun *ta* are homophones written as 他 and 她 respectively and the homophonic masculine and feminine reflexive *taziji* are written as 他自己 and 她自己 respectively<sup>3</sup>. This unique feature, although discussed in the literature, is often ignored by researchers because the Chinese examples in the linguistic literature are often transcribed using the Latin alphabet instead of the original script (Chinese characters). For example, Pan (1998) only mentions this orthographic distinction in a footnote without discussing any role this feature might play in the interpretations of Chinese anaphora. Some researchers have even proposed that Chinese speakers are not sensitive to this gender information at all and use the masculine form (i.e., pronoun 他 or reflexive 他自己) as the default form to refer both male and female referents (e.g., Su et al. 2016).

These two types of Chinese reflexives (*ziji* and *taziji*) have drawn much attention in the field as a critical test bed for evaluating both theoretical and psycholinguistic models of anaphora. For successful theoretical anaphor models, they should account for the existence of both types of reflexives in the same/different language(s) and their different behaviors as well: e.g., Reinhart and Reuland (1993) proposed two types of anaphoric elements cross-linguistically with SELF-anaphors as locally bound and SE-anaphors as non-local bound. Another similar approach is to argue that only the locally-bound reflexive is a true anaphor and the non-local ones are logophors which are subject to a different set of constraints (e.g., C.-T. J. Huang and Liu 2001). For the psycholinguistics field, these two types of reflexives have been widely used, beyond English (e.g., see a review in Jäger, Engelmann, and Vasishth 2015), to adjudicate competing claims (e.g., structure first vs. multiple constraint approaches), to evaluate predominant processing models (e.g.,

<sup>2</sup>We only focus on Mandarin Chinese in this study. For many variants/dialects of Chinese, they only have one form for pronouns: e.g., Cantonese only has one pronoun *keoi* and Hokkien only has one pronoun 伊 *yi* for both masculine and feminine forms.

<sup>3</sup>We consider it a pure orthographic (rather than morphological) feature because the homophonic characters 他 and 她 only differ in their semantic radicals (亻 vs. 女 and both share the same phonetic radical 也) and these radicals are not morphemes (see a comprehensive review of the Chinese writing system in Shu 2003).

cue-based retrieval models), and to investigate how both structural (e.g., locality and c-command) and non-structural constraints (e.g., gender or animacy cues) affect anaphor resolutions (e.g., Chen, Jäger, and Vasishth 2012; Dillon et al. 2014; Dillon, Chow, and Xiang 2016; Chang et al. 2020). However, despite the research on these forms, some basic behaviors of these reflexives still remain controversial. In this study, we focus on one of them which is rarely discussed in the literature: non-local bindings for *taziji*.

### 1.1 Long-distance binding options for *taziji*

Although most of the literature, e.g., C.-T. J. Huang, Li, and Li (2009), assumes that Chinese *taziji* is similar to English himself/herself in strictly requiring a local antecedent, there are still some exceptional cases reported in the literature. Most of these cases can be categorized as due to the influence of other (non-syntactic) factors. For example, Pan and colleague argued that locality is not a strict condition and both prominence (e.g., *animacy hierarchy*) and locality regulate the interpretation of *taziji* (Pan 1998; Pan and Hu 2003).

Pan (1998) claimed that in (2), where the local subject is inanimate, *taziji* can only take *Zhangsan* but not the local subject *book* as the antecedent following a “Prominence Constraint” that the animate subject is more “prominent” than the inanimate one in the animacy hierarchy.

(2) Pan (1998): Example 4.

张三 说 那 本 书 害 了 他 自 己  
 Zhangsan<sub>i</sub> shuo na ben shu<sub>j</sub> hai-le taziji<sub>i/\*j</sub>  
 Zhangsan<sub>i</sub> say that CL book<sub>j</sub> hurt-Perf himself<sub>i/\*j</sub>  
 ‘Zhangsan said that that book hurt himself.’

Yu (1992) also provided some data challenging the strict local binding requirement of *taziji*. For instance, if the local antecedent mismatches the person feature of *taziji*, as in (3), only the non-local subject John but not *wo* “I” can be the antecedent.

(3) Yu (1992): Example 9a.

约翰 要 我 坐 在 他 自 己 的 身 边。  
 Yuehan<sub>i</sub> yao wo<sub>j</sub> zuo zai taziji<sub>i/\*j</sub> de shenbian.  
 John<sub>i</sub> want I<sub>j</sub> sit at himself<sub>i/\*j</sub> ’s side.  
 ‘John wants me to sit at his side.’

Similarly, he argued for a “gender match” requirement for *taziji* as well. In the following examples, *Zhangsan* and *Lisi* are typical male names and *Weiling* is a typical female name in Chinese. According to Yu (1992), although marked as ? in (4), both *Zhangsan* and *Lisi* can be the antecedent of masculine *taziji* (他自己) while in (5) only the non-local *Zhangsan* can be the antecedent because *Zhangsan* is a typical male name but *Weiling* is not. This is what Pan and Hu (2003) categorized as a “Feature Compatibility Constraint” which requires that two coindexed elements must have compatible features (e.g., person or gender feature).

(4) Yu (1992): Example 9e.

?张三 叫 李四 吹捧 他自己  
 Zhangsan<sub>i</sub> jiao Lisi<sub>j</sub> chuipeng taziji<sub>i/j</sub>  
 Zhangsan<sub>i</sub> ask Lisi<sub>j</sub> promote himself<sub>i/j</sub>  
 ‘?Zhangsan asked Lisi to promote him/himself.’

(5) Yu (1992): Example 10.

张三 叫 魏玲 吹捧 他自己。  
 Zhangsan<sub>i</sub> jiao Weiling<sub>j</sub> chuipeng taziji<sub>i/\*j</sub>.  
 Zhangsan<sub>i</sub> ask Weiling<sub>j</sub> promote himself<sub>i/\*j</sub>.  
 ‘Zhangsan asked Weiling to promote him.’

Another line of unexpected cases comes from experimental studies on Chinese reflexives processing. Although most of them focus on the bare reflexive *ziji* (e.g., Dillon et al. 2014; Chen, Jäger, and Vasishth 2012; also see a comprehensive review in Jäger, Engelmann, and Vasishth 2015), there are a few studies looking at “long-distance” options for *taziji* as well. For instance, Dillon, Chow, and Xiang (2016) found a less-than-expected locality effect for *taziji* with a subcommanding antecedent (i.e., subject antecedents properly contained within a c-commander of *taziji*). In an eye-tracking-during reading-experiment, Chang et al. (2020) found no (gender-matching) interference effect of a long-distance antecedent but the off-line judgment questionnaire did show a significant amount of selections of the long-distance antecedent (~20%). The authors interpreted this as an “error” due to the experimental design and processing pressures. This long-distance binding was also found using other off-line tasks as well, e.g., Lyu and Kaiser (2021). However, these studies did not directly focusing on *taziji* (i.e., using it as a baseline comparison to *ziji*) and only provide indirect evidence for long-distance *taziji* or even treated them as a non-linguistic effect (e.g., processing errors).

Taken together, these cases of the long-distance binding option show that the antecedent of *taziji*, similar to English reflexives, should agree with the animacy, person and gender features of the reflexive, though the gender feature, for instance, is marked through the use of different characters (homophone *taziji*: 他自己 *himself* vs. 她自己 *herself*). *Taziji* can take a non-local antecedent when no matching one available locally. Also, as proposed by those scholars, there is also a possibility that the long-distance binding option is merely a by-product of other constraints or that the long-distance binding option is only allowed in some special syntactic environments. For example, in (2) the local subject is inanimate, so the long distance option is the only possibility given the animacy requirement<sup>4</sup> of *taziji* (他自己). Examples (3)–(5) all involve “control” structures, and although the binding theory also predicts a local binding for *taziji* in such structures, the fact that these are control structures complicates the interpretation. If both local and long-distance antecedents are exactly the same except for their relative positions in a sentence similar to (1), are both local and long-distance options still available for the reflexive *taziji*? If so, then what is the alternative explanation for those long-distance cases?

<sup>4</sup>Here we consider animacy as a semantic feature. It is also possible to treat animacy as a syntactic feature: e.g., Asudeh(1998) used animacy as a syntactic requirement in his HPSG account of *ziji*.

## 1.2 Proposal: long-distance binding option for *taziji* is inherently available

We propose that the long-distance binding option reflects the non-obligatory locality requirement for the reflexives *taziji*: the long-distance binding option of *taziji* is inherently available. Here we operationalize the concept of the “inherent long-distance” as: a non-local binding option is still available even when a local legitimate antecedent exists. For instance, in the example (1) above, we would show that *taziji* can also take the matrix subject *Zhangsan* as its antecedent even though the local subject *Lisi* is also a legitimate one. Although this proposal is in apparent conflict with some previous accounts of *taziji*, we will demonstrate in our experiments, that the strict locality requirement claimed in the literature is too strong, i.e., that “locality” is a preference rather than a requirement for *taziji*.

A natural empirical test for this proposal is also straightforward: in a structure with both local and non-local antecedents, we will test if speakers have a long-distance interpretation of *taziji*, when a local interpretation is also available. To do so, we constructed a bi-clausal structure “*Name1-says-Name2-Verb-taziji*” in which two potential antecedents only differed in their relative positions (local vs. long-distance) in the sentence while other structural aspects were controlled, i.e., they are all proper names in subject positions. As discussed previously, the antecedent needs to match the gender of *taziji* as well. We manipulated this gender agreement feature using gender stereotyped names (for antecedents) and different gender *taziji* (through different orthographies). By manipulating both syntactic and gender factors, we were able to explore whether the long-distance binding option for *taziji* is available above and beyond the effect of other constraints (i.e., gender-matching constraints in this case).

The rest of the paper is organized as follows: we first demonstrate the validity of our method in Experiment 1 on the pronoun *ta*, showing that our method is able to reveal a robust pattern in which a clear structural constraint exists. Next in Experiment 2, we test our proposal of the inherent long-distance (LD) option for *taziji* that both local and non-local bindings are available at the same time. In Experiment 3 and 4, we refute a hypothesis of the logophoric interpretation of the long-distance *taziji*. Last, we also discuss the implications of this study.

## 2 Experiment 1 & 2: inherent LD options for *taziji*

### 2.1 Overview

Since Experiment 1 and 2 share a similar design, items and task, we will first describe the general method, then present results separately for each experiment. Adopting the same sentence structure across both experiments (viz. *Name1-V1-Name2-V2-anaphor*), Experiment 1 serves as a control study and sanity check to demonstrate the validity of our method. Then in Experiment 2, we tested the hypothesis of inherent long-distance binding options for the reflexive *taziji*. The materials, data and code for this study are available in the OSF project repository (<https://shorturl.at/ikQV8>).

### 2.2 Methods

#### 2.2.1 Participants

84 native Chinese speakers (28 women; 28.64 +/- 5.99 years) from Mainland China were recruited through the online platform “Witmart” which is the largest China-based crowd-sourcing pool. This

platform has been used widely for social science research (see To and Lai 2015 for a review) and particularly linguistics-related studies on the Chinese population (e.g., Zhan, Levy, and Kehler 2020). Participants were randomly assigned to one of the experiments (42 participants for each experiment) and paid about 2 US dollars for participation.

### 2.2.2 Design and materials

Bi-clausal sentences of the structure “*Name1-V1-Name2-V2-pronoun/reflexive*” were designed for the critical items. For clarity and consistency purposes, two potential antecedents (i.e., Name 1 and 2) were labeled as the “Target” or “Competitor” with respect to the predictions of Binding principles: if the position of the antecedent was consistent with what Binding principles indicated, we then labeled it as the “Target”, otherwise we labeled it as the “Competitor”. For example, in the reflexive conditions, the local subject “Name2” was labeled as the “Target” and the subject of the matrix clause Name1 was labeled as “Competitor”. This enabled us to evaluate the relevant strength of the structural constraints (i.e., binding principles) within and between pronouns and reflexives. The “V1” was one of a set of verbs like “say” that select for a sentential complement while the “V2” was taken from a set of transitive verbs that were normed in a previous study (Y. Xu and Runner 2019) to ensure that both local and long-distance binding options were plausible. We provide a detailed description of these norming processes in the section 2.2.3 .

For both experiments, three factors were manipulated: the first two manipulations were the gender type of the *Target* as well as the *Competitor* (either matched or mismatched the gender of the pro-form) using gender stereotyped names taken from a previous study (Qiu et al. 2012). By manipulating the “gender” feature for both antecedents, we were able to test the separate effects of two factors (locality vs. gender matching) as well as their relative weights. Crucially, it enabled us to evaluate whether or not speakers chose the long-distance antecedents even when a local legitimate antecedent was available. The third manipulation was the gender of the pro-form (masculine or feminine form using different characters: 他 vs. 她; 他自己 vs. 她自己). Although this manipulation was not relevant to our main research question<sup>5</sup>, from a design perspective, the full factorial design has the advantage of achieving a balance of the stereotyped names in Name 1 and 2 positions.

A set of sample materials is shown in Table 1. A total of 16 sets of critical items and 40 filler sentences were created. We also included one of C.-T. J. Huang, Li, and Li (2009)’s original examples (as shown in 1a & b) at the beginning of each list as a direct replication to see if there is any effect of different lexical items used in our experimental stimuli. Recent studies demonstrated the importance of direct replications of original examples in evaluating the reliability of linguistic judgments (e.g., Chen, Xu, and Xie 2020). In total, each participant read 57 sentences for the experiment.

### 2.2.3 Norming of the materials

For gender stereotyped names (Name 1 & 2), we selected 32 male and 32 female Chinese names based on the pre-normed results from Qiu et al. (2012). Those names had been rated on a 5-point scale with 1 being extreme feminine and 5 being extreme masculine. The average gender

<sup>5</sup>However, as we mentioned in the introduction, a few researchers have claimed the masculine form could be used as the default form to refer to both male and female referents (e.g., Su et al. 2016). This manipulation gives us a chance to test this claim (which is not true according to our results).

stereotyped rating was 4.69 +/- 0.10 for male names and 1.16 +/- 0.07 for female names. The ratings differences for male and female names were significant ( $t(62) = -165.15, p < .0001$ ). For the same gender name pairs, the ratings of Name 1 positions were not different from those of Name 2 positions (male groups,  $t(30) = -0.001, p = .99$ ; female groups,  $t(30) = -0.04, p = .97$ ).

For verbs in the matrix clause (V1), Y. Xu and Runner (2019) selected 9 verbs such as *renwei* “think” or *biaoshi* “say” which can naturally select a sentential complement. Those verbs were taken from previous experimental studies on Chinese anaphor resolution by Chen, Jäger, and Vasishth (2012) and Dillon et al. (2014).

For verbs in the embedded clause (V2), they were taken from a previous studies by Y. Xu and Runner (2019). In order to ensure that both local and long distance binding interpretations were plausible, Y. Xu and Runner (2019) performed a separate norming task where 31 additional native Chinese speakers participated through a web-based questionnaire. Following the norming method introduced by Li and Zhou (2010) and Schumacher, Bisang, and Sun (2011), they created 45 sentence completion tasks in the form of “Name (e.g., *Zhangsan*) + *zai* (aspect marker) + Verb \_\_\_”. The choices were A. *ziji* (reflexive), B. *a proper name* (e.g., *Zhangsan*) and C. *Both*. If people chose the “C. *Both*”, then they were asked to judge from a 7-point scale to show which end (*ziji* or *Zhangsan*) they thought to be more plausible for the current sentence. For the finally chosen 16 “neutral” verbs, at least 26 participants (out of 31) chose to be ambiguous (i.e., choice of the C. *Both* with an average score of 4.21 +/- 1.06 on a 7-point scale).

Table 1: Sample materials and tasks used in Experiment 1 & 2.

<b>Sample materials</b>					
<b>Experiment 1: pronoun ta</b>					
孙志/张艳	表示	国俊/陈凤	在	低估	他/她
Sun Zhi/Zhang Yan	biaoshi	Guo Jun/Chen Feng	zai	digu	ta/ta
male/female name	say	male/female name	PROG	underestimate	him/her
'Name 1 says that Name 2 is underestimating him/her.'					
<b>Experiment 2: reflexive taziji</b>					
孙志/张艳	表示	国俊/陈凤	在	低估	他自己/她自己
Sun Zhi/Zhang Yan	biaoshi	Guo Jun/Chen Feng	zai	digu	taziji/taziji
male/female name	say	male/female name	PROG	underestimate	himself/herself
'Name 1 says that Name 2 is underestimating himself/herself.'					
<b>Tasks</b>					
*Antecedent choice task:			*Acceptability judgment task:		
句子里的他/他自己指的是谁?			你觉得这句话读起来如何?		
Who does <i>ta/taziji</i> refer to?			How does this sentence sound to you?		
(Name 1 / Name 2)			(7-point scale: very bad : very good)		

## 2.2.4 Procedure

The experiments were performed using the online questionnaire platform *Qualtrics* with each participant randomly assigned to one of the 8 experimental lists for each experiment. Critical

items were distributed across the lists following a Latin square design and were interspersed pseudo-randomly among filler sentences such that no two critical items were adjacent. One sentence was presented to the participants each time with two questions. For critical items, the first was an antecedent choice task (e.g., “who does *taziji* refer to?”) with two names mentioned in the sentence as choices (*Target* and *Competitor*). The second was an acceptability judgment task (“How does this sentence sound to you?”) to solicit participants’ ratings of each sentence using a 7-point Likert scale (1: very bad, 7: very good). For filler trials, similar comprehension and acceptability judgment questions were used. Before test trials, a practice session was included to help participants understand the task. The original example from C.-T. J. Huang, Li, and Li (2009) was always presented at the beginning of each experimental list. It took approximately 15 to 20 minutes to complete the task.

### 2.2.5 Analytical strategy

For statistical analysis, logistic (for antecedent choice data) or linear (for acceptability judgment data) mixed effects models were used with fixed effects of the *Target* gender type (match: 0.5; mismatch: -0.5), the *Competitor* gender type (match: 0.5; mismatch: -0.5) and the pro-form gender type (masculine: 0.5; feminine: -0.5). Following Barr et al. (2013), we also include the maximal random effects structure for both subjects and items in all models. When the maximal model did not converge, the random component with the least variance was removed and the model was refit until it converged. The dependent variable was choice of the Target (1: chosen; 0: not chosen) for antecedent choice data and the z-score transformed judgment score for acceptability judgment data. The z-score transformation was applied to eliminate potential scale bias which is a common strategy in the field (Schütze and Sprouse 2013). We also modeled the raw acceptability rating data as recommended by some researchers (e.g., Juzek 2015) which yielded similar results as z-score models. For clarity, we only present the z-scores analyses and plots here. Readers can refer to the R scripts available in the OSF repository for details.

## 2.3 Experiment 1 results: pronoun *ta*

### 2.3.1 Results summary

Figure 1 present data from Experiment 1 for both antecedent choice and acceptability judgment tasks. For antecedent choice data, we plotted the choice proportions for both Target (dark bars) and Competitor (light bars) across Target gender conditions (match/mismatch: left/right panels) and Competitor gender conditions (match/mismatch: left/right side for each panel). For acceptability judgment data, we plotted ratings across both Target gender conditions (match/mismatch: solid/dash lines) and Competitor gender conditions (match/mismatch: left and right sides).

The antecedent choice data patterns showed a clear structural constraint such that participants had overwhelming more Target choices across all conditions and their judgment ratings for Target match conditions were higher than mismatched ones: we observed a main effect of the Target gender type such that people had more Target choices when the Target matched the gender of the pronoun than when it mismatched ( $\hat{\beta} = 5.83, SE = 2.59, z = 2.25, p = .02$ ). Also, the ratings for Target gender matched ones were significantly higher than mismatched ones ( $\hat{\beta} = 0.52, SE = 0.07, t = 7.06, p < .001$ ). No other effects were observed.

Next, we focused on the Target gender match conditions (Figure 1: Left panel of the antecedent choice plot) to determine whether or not participants also considered the BT-incompatible choice (i.e., Competitor) when there was an acceptable BT-compatible antecedent available (i.e., gender match Target). First, following the standard proposed by He and Kaiser (2016), we compared the Competitor choices against 0% when both Target and Competitor matched the pronoun gender: the choice of the Competitor (average 3.1%) was not significantly higher than 0% (using log-transformed individual proportion means:  $\hat{\beta} = -0.05, SE = 0.04, t = 1.37, p = .18$ ). Second, we tested the Competitor choices between the Competitor match and mismatch conditions. Comparing to the baseline Competitor gender mismatch condition, participants did not choose more Competitors even when they matched the gender of the pronoun *ta* ( $\hat{\beta} = -0.61, SE = 5.23, z = -0.12, p = .91$ ). Both analyses indicate that participants did not consider the Competitor as an acceptable antecedent for the pronoun *ta*.

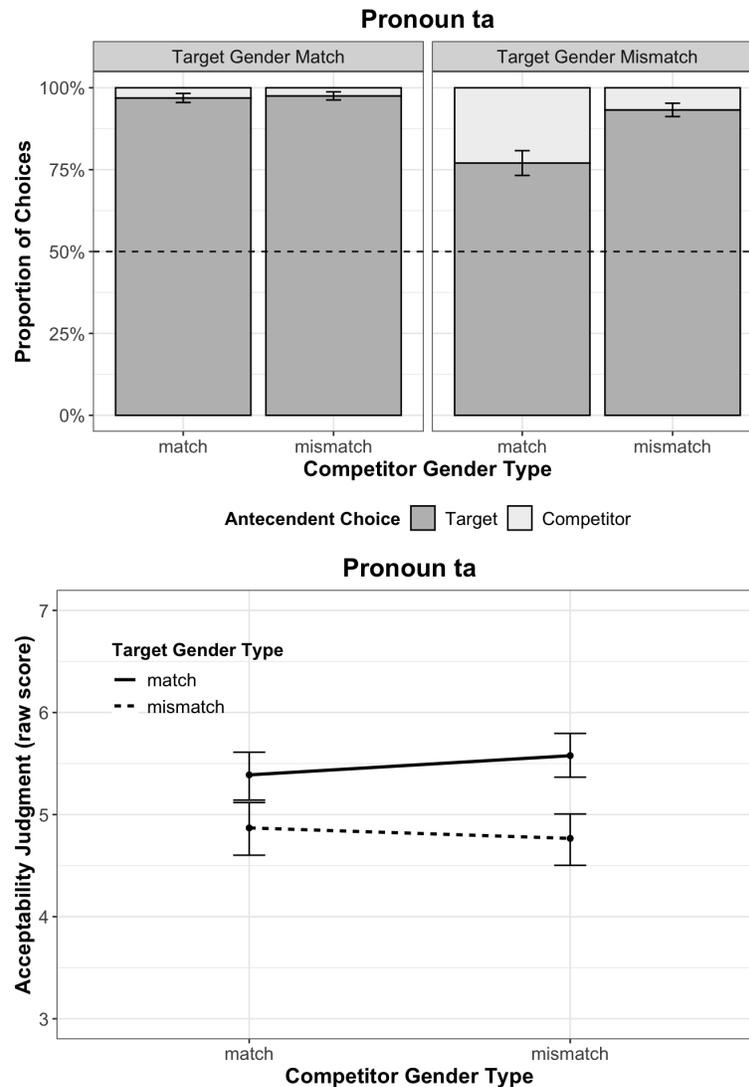


Figure 1: Antecedent choice and acceptability judgment (z-score transformed) results in Experiment 1. For choice results, dark bars represent the proportion of choosing the Target/Name1 as the antecedent of the pronoun *ta*, while the light bars represent the proportion of choosing the Competitor/Name2 as the antecedent. Error bars represent the standard error of the subject means. The judgment results figure shows the averaged z-score transformed judgment ratings across subjects and items in each condition. Solid lines represent conditions where the Target/Name1 matched the gender of the pronoun *ta* while dash lines are for the conditions where the Target mismatched the gender of the pronoun. Error bars represent bootstrapped 95% confidence intervals.

### 2.3.2 Predictions for Experiment 2/*taziji* results

Experiment 1 demonstrated the validity of our method. Participants were sensitive to both our gender manipulation as well as the structural constraints: while overwhelmingly they selected binding theory compatible antecedents across all conditions, they provided lower acceptability ratings for Target gender mismatch conditions. Thus, our methods are able to reveal a robust pattern when there is indeed a strong structural constraint: participants would only select BT-compatible antecedents as shown in their antecedent choice task and provided lower ratings when

the Target mismatched the reflexive gender. If the reflexive *taziji* also follows a strict syntactic constraint, we should observe a similar effect as that for the pronoun *ta* (i.e., overwhelmingly Target choices across all conditions). If instead *taziji* has an inherent long-distance binding option, we should see significant choices of the structural inaccessible but gender matching Competitor even when a gender match Target available. Next, we present results from Experiment 2 showing the long-distance binding option for the reflexive *taziji*.

## 2.4 Experiment 2 results: reflexive *taziji*

### 2.4.1 Results summary

Similar to Experiment 1, Figure 2 presents data from Experiment 2 for both antecedent choice and acceptability judgment tasks. For reflexive conditions, we observed a main effect of the target gender type ( $\hat{\beta} = 2.49, SE = 0.71, z = 3.53, p < .001$ ) and of the competitor gender type ( $\hat{\beta} = -2.36, SE = 0.71, z = -3.31, p < .001$ ) in people's antecedent choices: people had more Target choices when the Target matched and also the Competitor mismatched the gender<sup>6</sup>. For the acceptability judgment task, we observed a main effect of the Target gender type ( $\hat{\beta} = 0.23, SE = 0.07, t = 3.03, p < .001$ ) as well as a trending interaction of the Target gender type and Competitor gender type ( $\hat{\beta} = -0.28, SE = 0.15, t = -1.89, p = .06$ ). We further conducted a planned pairwise comparison showing that the ratings for the Target matched vs. mismatched conditions did not differ under the Competitor gender matched conditions ( $\hat{\beta} = 0.17, SE = 0.20, t = 0.83, p = .41$ ) and only differed under Competitor gender mismatched conditions ( $\hat{\beta} = 0.65, SE = 0.20, t = 3.22, p < .001$ ): people rated sentences with only the Competitor gender match as perfectly legitimate showing no significant difference from the ones in which the Target gender also matched ones.

Similar to Experiment 1, we also compared the Competitor choices under the Target gender match conditions (i.e., Figure 2: Left panel of the antecedent choice data) to test if participants also considered the binding theory incompatible but gender matched Competitor as the antecedent of the reflexive *taziji*. First, we compared the Competitor choices when both Target and Competitor matched the gender: the choice of the Competitor (average 19.4%) was significantly higher than 0% (using log-transformed individual proportion means:  $\hat{\beta} = 0.09, SE = 0.04, t = 2.36, p < .03$ ). Second, we found people had significantly more Competitor choices when the Competitor also matched the condition ( $\hat{\beta} = 2.79, SE = 1.28, z = 2.17, p < .03$ ) comparing to the Target-only matched conditions. Both results indicate that people consider both the Target and Competitor as the potential antecedents of the reflexive *taziji*.

<sup>6</sup>We also observed a main effect of the reflexive gender such that people had more Target choices under masculine conditions (他自己 *himself*) than feminine conditions (她自己 *herself*). One possibility is that people treated the masculine form as the default and used it to refer to a feminine antecedent as well. Another possibility is that the masculine form is much frequent (almost six times more) than the feminine one (estimation based on a Google Ngram search from 1988 to 2008) and the uncertainty was reduced (i.e., fewer Competitor choices) when a more frequent form was presented. The interpretation of this effect won't change the main conclusions of this study and we leave this as an open question and expect future work to address this phenomenon.

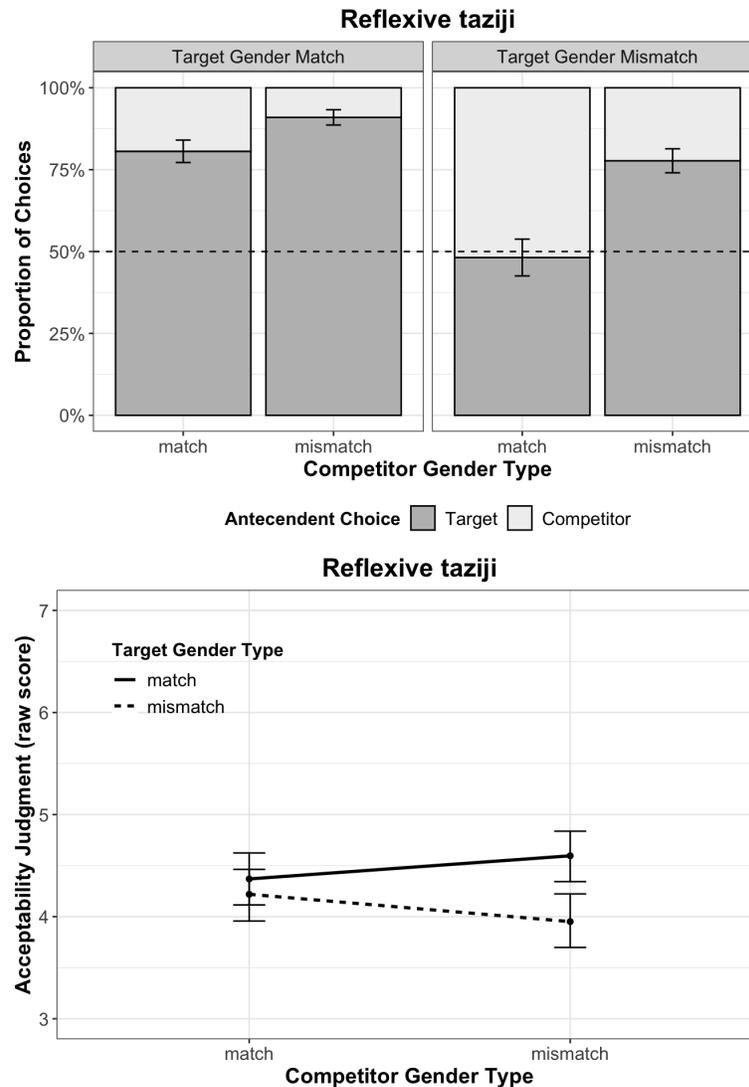


Figure 2: Antecedent choice and acceptability judgment (z-score transformed) results in Experiment 2. For choice results, dark bars represent the proportion of choosing the Target/Name2 as the antecedent of the reflexive *taziji*, while the light bars represent the proportion of choosing the Competitor/Name1 as the antecedent. Error bars represent the standard error of the subject means. The judgment results figure shows the averaged z-score transformed judgment ratings across subjects and items in each condition. Solid lines represent conditions where the Target/Name2 matched the gender of the reflexive *taziji* while dash lines are for the conditions where the Target mismatched the gender of the reflexive. Error bars represent bootstrapped 95% confidence intervals.

#### 2.4.2 *Ta* vs. *taziji* analysis and textbook original examples

We conducted a “cross-experiment” analysis using similar mixed-effects models of the antecedent choices, as well as sentences acceptability ratings, comparing the pronoun *ta* and the reflexive *taziji* (treatment coded, *ta*: 0; *taziji*: 1) collapsed across all conditions. Results showed that participants had more Target choices (i.e., BT-compatible choices) for pronoun conditions than reflexive conditions ( $\hat{\beta} = -1.57, SE = 0.33, z = -4.75, p < .0001$ ) and the acceptability ratings for pronoun conditions did not differ from reflexive conditions ( $\hat{\beta} = 0.01, SE = 0.05, t = 0.01, p > .9$ ).

Compared to the pronoun *ta*, the reflexive *taziji* was less sensitive to the structural constraints such that both local and non-local subjects were considered as valid antecedents.

We also tested the original examples of the pronoun *ta* and the reflexive *taziji* (i.e., 1a and 1b) used in C.-T. J. Huang, Li, and Li (2009) textbook. Because they used two same-gender names (i.e., *Zhangsan* or *Lisi*) without any gender manipulation in their original examples, the corresponding conditions in our experiment should be Target gender-matched with Competitor gender-matched ones. Participants had 95.2% Target and 4.8% Competitor choices for the *ta* example and 69.8% Target and 30.2% Competitor choices for the *taziji* example. The Target choices for the *ta* example was significantly higher than those for the *taziji* example ( $\hat{\beta} = -2.16, SE = 0.80, z = -2.71, p < .01$ ). These patterns mirrored our experimental data shown in the *ta* and *taziji* experiment.

### 2.4.3 Intermediate summary

In Experiment 1, we demonstrated the validity of our method which revealed a robust pattern where the syntactic constraint served a major role in the interpretation of the pronoun *ta*. In Experiment 2, with the same structure and lexical items, the reflexive *taziji* showed a significant amount of non-local antecedent choices even when a local gender matching antecedent was available. The results of C.-T. J. Huang, Li, and Li (2009)'s original examples were consistent with the corresponding conditions in our experimental trials, which suggests that the patterns in our experimental data were not driven by the particular lexical items used in our experiments. Although the experiments themselves already provided sufficient evidence for an inherent long-distance binding option for *taziji*, there are still alternative explanations. For example, the use of the matrix verb like “say” in Experiment 2 could introduce a “source of information” or “logophoric center” and this logophoricity interpretation could arguably be an important factor responsible for the non-local bindings observed cross-linguistically (see a review: Charnavel et al. 2017). Thus, we shall explore this possibility in Experiments 3 & 4.

## 3 Experiment 3 & 4: LD *taziji* is not driven by logophoricity

### 3.1 Overview

In Experiment 3 & 4, we aim to assess the potential confounding factors of logophoricity when interpreting the reflexive *taziji*. As shown in Experiment 2, we used verbs like “say” or “think” in the matrix clause in order to introduce a clausal complement. Such verbs could potentially introduce a “source of information” (i.e., the long-distance antecedent/Name1 in our case) which is arguably one of the dominant discourse factors constraining logophors (Charnavel et al. 2017), for instance, in the English “picture noun phrase” structure (e.g., Pollard and Sag 1992 ) and the Chinese logophor *ziji* (e.g., C.-T. J. Huang and Liu 2001). Following this view, the long-distance binding we observed for *taziji* could be driven by an independent effect of the logophoric interpretations.

Unfortunately, there is no uncontroversial diagnostic tool available to verify logophoricity directly (Charnavel et al. 2017); instead our approach here is to exclude this possibility by showing that the long-distance *taziji* observed previously was *not* driven by any logophoric effects. If the “logophor hypothesis” is correct, we should expect the choices of the long-distance *taziji* would be modulated by the logophoric conditions: more long-distance antecedents should be selected when

they are the logophoric center (e.g., source of information, attitude holder, etc.) than when they are not. Alternatively, if the long-distance *taziji* is not driven by logophoricity, we should not see any difference in long-distance interpretations for *taziji* whether it is a logophoric center or not.

In these experiments, adopting the same structure in Experiments 1 & 2, we manipulated the matrix verbs to create a “source vs. perceiver” contrast for long-distance antecedents (see Table 2 for an example) and investigate how people interpreted the reflexive *taziji* (Experiment 3) and *ziji* (Experiment 4). We also include the *ziji* as a baseline comparison and to see how different these two types of reflexives are with respect to the sensitivity to the logophoric manipulations.

## 3.2 Methods

### 3.2.1 Participants and procedure

Another 94 native Mandarin speakers (64 women; 24.13 +/- 6.74 years) were recruited from the university communities in Nanjing, China. Participants were randomly assigned to one of the experiments (50 for Experiment 3 and 44 for Experiment 4) and paid 10 Chinese Yuan (~1.6 US dollars) for participation. The procedure was similar to Experiment 1 and 2 and implemented on the questionnaire platform *Qualtrics*.

### 3.2.2 Design and materials

A similar bi-clausal structure (*Name1-think/hear-Name2-V2-taziji/ziji*) with two same-gender names and gender-matching reflexives were used to ask participants to select the antecedents and rate the sentences. We used the same set of names and embedded verbs as in Experiments 1 and 2. We manipulated the logophoricity center using the “*renweilthink*” vs. “*tingshuohear*” contrast in the matrix verb position. We chose this contrast pair because the verb “think” creates a stronger logophoric center because it indicates both a “source of information” and an “attitude holder” (Charnavel et al. 2017), and the verb “hear” clearly indicates the opposite of a source as the “perceiver of information” [Kaiser et al. (2009)]<sup>7</sup>. In total, 16 sets of critical items and 40 filler sentences were used. A set of sample materials are shown in Table 2.

<sup>7</sup>In a pilot study, we also used the “*biaoshisay*” vs. “*tingshuohear*” pair as the logophoricity manipulation. The results also showed a similar pattern as what we found in this experiment.

Table 2: Sample materials and tasks used in Experiment 3 &amp; 4.

Sample materials					
<i>Experiment 3: reflexive taziji</i>					
孙志	认为/听说	国俊	在	低估	他自己
Sun Zhi	renwei/tingshuo	Guo Jun	zai	digu	taziji
male name	think/hear	male name	PROG	underestimate	himself
'Name 1 thinks/hears that Name 2 is underestimating himself.'					
<i>Experiment 4: reflexive ziji</i>					
孙志	认为/听说	国俊	在	低估	自己
Sun Zhi	renwei/tingshuo	Guo Jun	zai	digu	ziji
male/female name	think/hear	male name	PROG	underestimate	self
'Name 1 thinks/hears that Name 2 is underestimating self.'					
Tasks					
*Antecedent choice task:			*Acceptability judgment task:		
句子里的他自己/自己指的是谁?			你觉得这句话读起来如何?		
Who does taziji/ziji refer to?			How does this sentence sound to you?		
(Name 1 / Name 2)			(7-point scale: very bad : very good)		

### 3.3 Results and discussions

Figure 3 presents the antecedent choice and acceptability judgment results of Experiment 3 and 4. For *taziji*, consistent with the pattern in Experiment 2, we found about 17.1% long-distance antecedent choices across conditions. However, no effect of the logophoricity manipulation was observed: there was no difference in long-distance choices between the “think” and “hear” conditions ( $\hat{\beta} = 0.07, SE = 0.21, z = 0.31, p = .76$ ). For *ziji*, the antecedent choices results were consistent with the claims in the literature for a typical long-distance reflexive that can take both local and non-local antecedents (~64% vs. 36%). Also as expected, it showed significant more long-distance choices than *taziji* (64% vs. 17%,  $\hat{\beta} = 2.19, SE = 0.12, z = 17.87, p < .0001$ ), but there was no difference between “think” and “hear” conditions under *ziji* either ( $\hat{\beta} = 0.14, SE = 0.18, z = 0.79, p = .43$ ). Lastly, the acceptability ratings did not differ across conditions and anaphors ( $ps > .05$ ).

The results supported our claim that the long-distance *taziji* is not a logophor or any by-product of logophoric effects (e.g., *source of information effect*): no difference in choices of the long-distance antecedent between the source/“think” and the perceiver/“hear” condition. Also, the lack of logophoric effects for *ziji* provided additional evidence against the “logophoricity hypothesis” in such environments (e.g., Pollard and Xue 2000 also argued that “source of information” does not license long distance *ziji*). *Taziji* is still different from *ziji* though in terms of the binding behavior: there was a clear long-distance preference for *ziji* and local preference for *taziji*. However, this difference is due to the different sensitivity to locality rather than logophoricity.

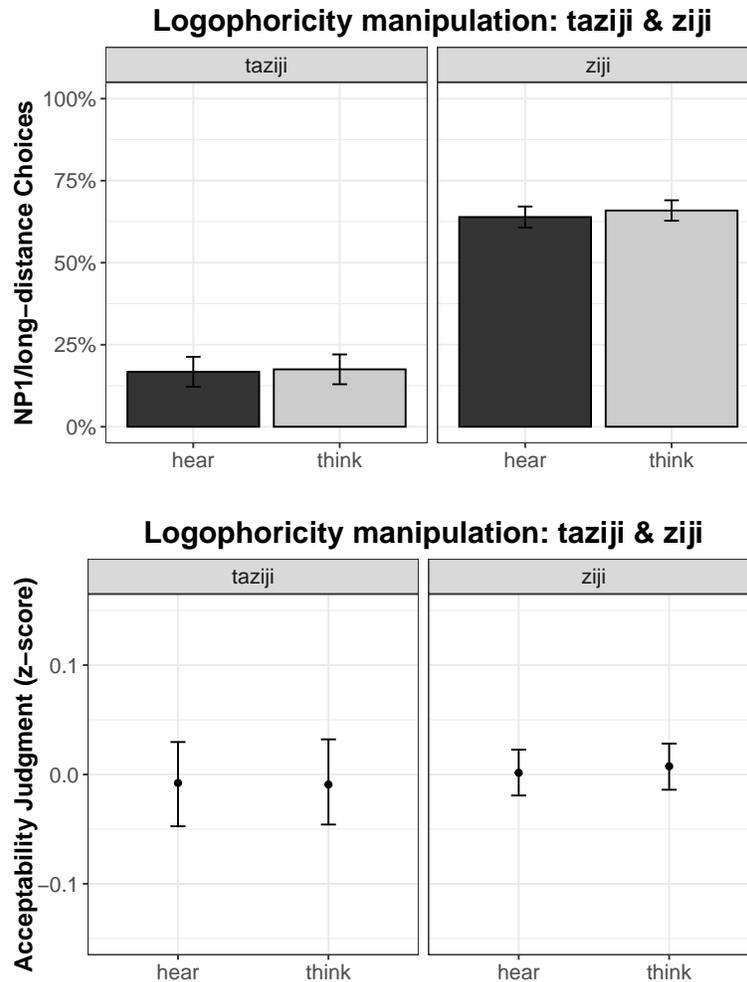


Figure 3: Long-distance Antecedent choice and acceptability judgment (z-score transformed) results for *taziji* (Experiment 3) and *ziji* (Experiment 4). For the choice results plot, dark and light bars represent the long-distance choices/Name1 under “hear” and “think” conditions respectively. The judgment results plot shows the averaged z-score transformed judgment ratings in each condition. Error bars represent the standard error of the subject means.

### 3.4 Summary of the experiments

Together, these experiments show that the reflexive *taziji* clearly exhibits a long-distance binding option even when a local legitimate antecedent is available. First, the pronoun *ta* conditions demonstrated the validity of our method and revealed the pattern when a strong syntactic constraint does have an effect on the interpretation of the anaphor. Using the pronoun *ta* as a baseline, we found a stronger structural constraint on pronoun conditions than reflexive conditions. For pronouns, people predominantly chose non-local subjects across all conditions. This illustrates a strong structural bias for a non-local antecedent, so strong that it can even override the gender matching requirements of the pronoun. Although people made fewer Target choices when the Target mismatched the gender of the pronoun *ta*, especially when the Competitor gender matched, their low rating indicated their sensitivity to the gender mismatch required in order to make that selection. On the other hand, the reflexive *taziji* showed a long-distance binding option even

when the Target also matched the gender of the reflexive: participants still had almost 20% of the non-local choice and the ratings did not differ from the Target only matched ones. The results of *The syntax of Chinese* textbook examples were consistent with the corresponding conditions in our experimental trials, which suggests that the patterns in our experimental data were not driven by the particular lexical items used in our experiments. Next, we refuted the “logophoricity hypothesis” arguing that the long-distance *taziji* is not a logophor or a by-product of the independent influence of a logophoric interpretation: the selections of the long-distance antecedent did not vary under the different logophoricity conditions. In short, the long-distance binding option is inherent for *taziji*.

## 4 General discussion

In this study, we explored the degree to which long-distance binding was available for the Mandarin Chinese reflexive *taziji*. We hypothesized that the non-local binding behavior hinted at in the previous literature was due to the fact that the reflexive *taziji* is inherently capable of long-distance binding. To the best of our knowledge, this is the first experimental study to demonstrate the inherent long-distance binding option for *taziji* together with a direct replication of textbook examples. Unlike phenomena where reasoning about syntactic analysis is based on sentence pairs illustrating a sharp contrast (i.e., acceptable vs. unacceptable), we believe that the long-distance binding option for *taziji* may be harder to detect without doing formal experiments: the local-binding option is still much preferred (about 80% of choices) for native speakers. Intuition alone may not be sensitive enough to distinguish whether the local-binding option is a requirement or preference. We do not aim to call into question the value of informal acceptability judgment methods in syntax research as there have been a number of studies demonstrating the reliability of syntax data collected using such methods (e.g., Sprouse and Almeida 2012; Sprouse, Schütze, and Almeida 2013). Instead, this case study aims to illustrate that formal experimental investigation can provide important insight not easily gained in other ways (see also Runner, Sussman, and Tanenhaus 2006). Next, we discuss some implications of our results from both theoretical and methodological perspectives.

### 4.1 Theoretical consequences

The long-distance binding option for the reflexive *taziji* immediately challenges current syntactic theories on the distribution of Chinese anaphora. Solutions of modifying binding theories like extending binding domains in Chinese may not work because the pronoun *ta* behaves well with respect to the defined binding domain: the antecedent of the pronoun *ta* must be outside of the binding domain (i.e., the local clause in our case). Alternatively, we could treat *taziji* as a kind of long-distance reflexive similar to *ziji*, but this in turn challenges some of the current views on long-distance reflexives too.

One of the features claimed to be relevant for long-distance reflexives is “monomorphemicity” (Pica 1987). For languages with both monomorphemic and polymorphemic reflexives (e.g., *ziji* and *taziji* in Mandarin Chinese), the claim has been that only the monomorphemic one can be long-distance bound while the polymorphemic is claimed to be strictly local (e.g., C.-T. J. Huang and Liu 2001). This is also argued to be a universal feature of long-distance reflexives cross-linguistically (e.g., Italian: Giorgi 1984; Norwegian: Hellan 1991; see a review in Charnavel et al. 2017). Also, the poly/monomorphemicity distinction in anaphor binding plays a key role in

some of the classic syntactic analyses of long-distance reflexives (e.g., Cole, Hermon, and Sung 1990; C.-T. J. Huang and Liu 2001) which proposed that the long-distance reflexive *ziji* undergoes successive-cyclic local movement in the LF component. Since only monomorphemic *ziji* is an  $X^0$  category which can undergo head-movement, as opposed to polymorphemic/phrasal *taziji*, we should only see the long-distance binding option for *ziji*, and not for *taziji*. Our data clearly call this claim into question.

Then what factors constrain the binding behavior of *taziji*? Does *taziji* follow the same constraints as those for *ziji*? (e.g., “subject orientation,” “blocking effects”; see a review in C.-T. J. Huang and Liu 2001). If so, why would a single language have two different types of long-distance reflexive? How do they differ? Cross-linguistically many languages have multiple reflexives, but are claimed to require a unique binding domain for each one: for example, Norwegian has three different reflexives *seg selv*, *seg* and *ham selv* and they differ with respect to their binding domains (see Dalrymple 1993 on LFG Binding Theories). This is clearly not the case for *taziji* and *ziji*: they share the same binding domain as shown in our results.

We suggest a promising first step is to conduct a series of studies directly comparing *ziji* and *taziji*. This may be necessary to tease apart their differences. Only after that can we know which empirical claims can be used to support our theoretical reasoning. One promising direction is to extend some non-uniform/mixed accounts on *ziji* to account for *taziji* as well (e.g., Pan 1998; Pollard and Sag 1992; C.-T. J. Huang and Liu 2001) or to account for both local and long-distance bindings using a different set of principles: see a recent attempt by Liu (2020) who proposed a unified logophoricity theory for both *ziji* and *taziji*.

## 4.2 Using formal experiments in developing linguistic theories

Here we do not aim to contribute to the recent debate on the reliability of data in the linguistic literature and whether or not linguistics—or syntax in particular—should employ more quantitative methods in developing theories (e.g., Sprouse and Almeida 2012; Gibson and Fedorenko 2013). Instead we want to suggest some advantages of using formal experiments in linguistic theorizing and offer a few tips in collecting judgment data and making the linguistic data more transparent and easier to share with peers.

First, using formal judgment experiments may help linguists attain reliable interpretations of subtle distinctions or controversial phenomena. Taking the current study as an example, although the long-distance binding option was clearly available for the reflexive *taziji*, the local binding option was preferred. This kind of distinction (preferred vs. obligatory), unlike other robust phenomena (e.g., blatant violation of word order), may not be easy to detect using informal means. We are not recommending doing formal experiments for every linguistic phenomenon, but we do believe they can be critical in revealing subtle effects not evident in informal acceptability judgments.

Also, performing formal experiments can help researchers tease apart different sources of constraints that influence a particular linguistic phenomenon. This can be useful for “interface” types of questions, for example, testing syntactic and semantic effects on anaphor resolution (e.g., Kaiser et al. 2009). Constructing sets of examples that vary on particular features in a controlled way can make it possible to tease apart the separate effects of each feature. In our case, we wanted to test the influence of both gender and structural constraints on people’s interpretations of the reflexive *taziji*, so we created four types of sentences manipulating the position of the antecedents

and their gender type. By doing so, we can see the influence of the syntactic and gender constraints (main effects) as well as their interaction (interaction effects) from the statistical analysis (see also Sprouse et al. 2016 for a similar approach on “island effects”).

Lastly, doing formal experiments can provide more transparency of the linguistic data used in the field. It is very common that for non-native linguists, the judgments of linguistic phenomenon heavily rely on the data provided by other linguists who are experts or native speakers of that language. Experimental results can inform researchers of the difference between two conditions (i.e., theory-driven contrast pairs used in linguistic theorizing) more than binary distinctions. Presenting raw data with statistical analysis help other researchers know more about a phenomena and evaluate the arguments proposed to analyze it<sup>8</sup>.

To recapitulate, contradicting the predominant claim in the field, we provide strong evidence from both formal experiments and replications of the textbook examples showing that the Chinese reflexive *taziji* has a long-distance binding option. Although further work is needed to fully characterize the binding behaviors of Mandarin Chinese reflexives in order to have a unified theory, we believe this study is a promising start and hope the experimental approach presented here will benefit the future studies.

## Open practices statement

The materials, data and code are available in the OSF project repository: <https://shorturl.at/ikQV8>

## References

- Asudeh, Arshia. 1998. “Anaphora and argument structure: Topics in the syntax and semantics of reflexives and reciprocals.” MPhil Dissertation, University of Edinburgh.
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. “Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal.” *Journal of Memory and Language* 68 (3): 255–78. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Chang, Wenshuo, Yunyan Duan, Jingjing Qian, Fuyun Wu, Xiaoming Jiang, and Xiaolin Zhou. 2020. “Gender Interference in Processing Chinese Compound Reflexive: Evidence from Reading Eye-Tracking.” *Language, Cognition and Neuroscience* 35 (10): 1355–70. <https://doi.org/10.1080/23273798.2020.1781213>.
- Charnavel, Isabelle, C.-T. James Huang, Peter Cole, and Gabriella Hermon. 2017. “Long-Distance Anaphora: Syntax and Discourse.” In *The Wiley Blackwell Companion to Syntax, Second Edition*, edited by Martin Everaert and Henk C. van Riemsdijk, 1–82. Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118358733.wbsyncom074>.
- Chen, Zhong, Lena Jäger, and Shravan Vasishth. 2012. “How Structure-Sensitive Is the Parser? Evidence from Mandarin Chinese.” In *Empirical Approaches to Linguistic Theory*, edited by Britta Stolterfoht and Sam Featherston, 43–62. DE GRUYTER. <https://doi.org/10.1515/9781614510888.43>.

---

<sup>8</sup>We also realize that there are some cases in which running experiments is extremely difficult, or not practical; for example, studies on endangered languages or contexts where native speakers are hard to reach. However, approaches have been proposed to address these limitations as well, for instance, the small N acceptability paradigm (Mahowald et al. 2016) to deal with the problem of limited participants.

- Chen, Zhong, Yuhang Xu, and Zhiguo Xie. 2020. "Assessing Introspective Linguistic Judgments Quantitatively: The Case of The Syntax of Chinese." *Journal of East Asian Linguistics* 29 (3): 311–36. <https://doi.org/10.1007/s10831-020-09210-y>.
- Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Studies in Generative Grammar [SGG]. De Gruyter. <https://books.google.com/books?id=XiJTlengwloC>.
- Cole, Peter, Gabriella Hermon, and Li-May Sung. 1990. "Principles and Parameters of Long-Distance Reflexives." *Linguistic Inquiry* 21 (1): 1–22. <http://www.jstor.org/stable/4178657>.
- Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. CSLI Publications.
- Dillon, Brian, Wing-Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu, and Colin Phillips. 2014. "The Structure-Sensitivity of Memory Access: Evidence from Mandarin Chinese." *Frontiers in Psychology* 5 (September). <https://doi.org/10.3389/fpsyg.2014.01025>.
- Dillon, Brian, Wing-Yee Chow, and Ming Xiang. 2016. "The Relationship Between Anaphor Features and Antecedent Retrieval: Comparing Mandarin Ziji and Ta-Ziji." *Frontiers in Psychology* 6 (January). <https://doi.org/10.3389/fpsyg.2015.01966>.
- Gibson, Edward, and Evelina Fedorenko. 2013. "The Need for Quantitative Methods in Syntax and Semantics Research." *Language and Cognitive Processes* 28 (1-2): 88–124. <https://doi.org/10.1080/01690965.2010.515080>.
- Giorgi, Alessandra. 1984. "Towards a Theory of Long Distance Anaphors: A GB Approach." *The Linguistic Review* 3 (4): 307–61. <https://doi.org/10.1515/tlir.1984.3.4.307>.
- He, Xiao, and Elsi Kaiser. 2016. "Processing the Chinese Reflexive "Ziji": Effects of Featural Constraints on Anaphor Resolution." *Frontiers in Psychology* 7 (April). <https://doi.org/10.3389/fpsyg.2016.00284>.
- Hellan, Lars. 1991. "Containment and Connectedness Anaphors." In *Long-Distance Anaphora*, edited by Jan Koster and Eric Reuland, 27–48. Cambridge, UK: Cambridge University Press.
- Hu, J. 2019. *Prominence and Locality in Grammar: The Syntax and Semantics of Wh-Questions and Reflexives*. Routledge Studies in Chinese Linguistics. Taylor & Francis.
- Huang, Cheng-Teh James, Yen-hui Audrey Li, and Yafei Li. 2009. *The Syntax of Chinese*. Cambridge, UK ; New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781139166935>.
- Huang, Cheng-Teh James, and Chen-Sheng Liu. 2001. "Logophoricity, Attitudes, and Ziji at the Interface." In *Between Syntax and Semantics*, 300–343. Routledge Taylor & Francis Group.
- Huang, C.-T. James, Yen-hui Audrey Li, and Yafei Li. 2013. *汉语句法学/The Syntax of Chinese*. Translated by Zhang Heyou and Gu Yang. Beijing, China: 世界图书出版公司.
- Huang, Yun-Hua. 1984. "Reflexives in Chinese." *Studies in English Literature and Linguistics* 10: 163–88.
- Jäger, Lena A., Felix Engelmann, and Shravan Vasishth. 2015. "Retrieval Interference in Reflexive Processing: Experimental Evidence from Mandarin, and Computational Modeling." *Frontiers in Psychology* 6 (May). <https://doi.org/10.3389/fpsyg.2015.00617>.
- Juzek, Tom. 2015. "Acceptability Judgement Tasks and Grammatical Theory." Doctoral dissertation, University of Oxford.
- Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman, and Michael K. Tanenhaus. 2009. "Structural and Semantic Constraints on the Resolution of Pronouns and Reflexives." *Cognition* 112 (1): 55–80. <https://doi.org/10.1016/j.cognition.2009.03.010>.

- Li, Xiaoqian, and Xiaolin Zhou. 2010. "Who Is Ziji? ERP Responses to the Chinese Reflexive Pronoun During Sentence Comprehension." *Brain Research* 1331 (May): 96–104. <https://doi.org/10.1016/j.brainres.2010.03.050>.
- Liu, Yingdong. 2020. "Logophoricity and Mandarin Exempt Reflexives." *University of Pennsylvania Working Papers in Linguistics* 26 (1): 18.
- Lyu, Jun, and Elsi Kaiser. 2021. "Antecedent Prominence and the Chinese Reflexive Ziji." Poster presented at the *34th Annual CUNY Conference on Human Sentence Processing*. <https://www.cuny2021.io/2021/02/24/279/>.
- Mahowald, Kyle, Peter Graff, Jeremy Hartman, and Edward Gibson. 2016. "SNAP Judgments: A Small N Acceptability Paradigm (SNAP) for Linguistic Acceptability Judgments." *Language* 92 (3): 619–35. <https://doi.org/10.1353/lan.2016.0052>.
- Pan, Haihua. 1998. "Closeness, Prominence, and Binding Theory." *Natural Language and Linguistic Theory* 16 (4): 771–815. <https://doi.org/10.1023/A:1006056414208>.
- Pan, Haihua, and Jianhua Hu. 2003. "Prominence and Locality in the Binding of Mandarin Complex Reflexive 'Ta-Ziji (s/He-Self).'" *Journal of Chinese Linguistics Monograph Series*, no. 19: 152–70. <http://www.jstor.org/stable/23825496>.
- Pica, Pierre. 1987. "On the Nature of the Reflexivization Cycle." In *Proceedings of the North East Linguistic Society*, 17:483–500. 2.
- Pollard, Carl, and Ivan A. Sag. 1992. "Anaphors in English and the Scope of Binding Theory." *Linguistic Inquiry* 23 (2): 261–303. <https://www.jstor.org/stable/4178768>.
- Pollard, Carl, and Ping Xue. 2000. "Syntactic And Nonsyntactic Constraints On Long-Distance Reflexives." In *Long Distance Reflexives*, edited by Peter Cole, Gabriella Hermon, and C.-T. James Huang, 317–42. BRILL. [https://doi.org/10.1163/9781849508742\\_011](https://doi.org/10.1163/9781849508742_011).
- Qiu, Lijing, Tamara Y. Swaab, Hsuan-Chih Chen, and Suiping Wang. 2012. "The Role of Gender Information in Pronoun Resolution: Evidence from Chinese." Edited by Kevin Paterson. *PLoS ONE* 7 (5): e36156. <https://doi.org/10.1371/journal.pone.0036156>.
- Reinhart, Tanya, and Eric Reuland. 1993. "Reflexivity." *Linguistic Inquiry* 24 (4): 657–720.
- Runner, Jeffrey T., Rachel S. Sussman, and Michael K. Tanenhaus. 2006. "Processing Reflexives and Pronouns in Picture Noun Phrase." *Cognitive Science* 30 (2): 193–241. [https://doi.org/10.1207/s15516709cog0000\\_58](https://doi.org/10.1207/s15516709cog0000_58).
- Schumacher, Petra B., Walter Bisang, and Linlin Sun. 2011. "Perspective in the Processing of the Chinese Reflexive Ziji: ERP Evidence." In *Anaphora Processing and Applications*, edited by Iris Hendrickx, Sobha Lalitha Devi, António Branco, and Ruslan Mitkov, 119–31. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Schütze, Carson T, and Jon Sprouse. 2013. "Judgment Data." In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 27–50.
- Shu, Hua. 2003. "Chinese Writing System and Learning to Read." *International Journal of Psychology* 38 (5): 274–85. <https://doi.org/10.1080/00207590344000060>.
- Sprouse, Jon, and Diogo Almeida. 2012. "Assessing the Reliability of Textbook Data in Syntax: Adger's *Core Syntax*." *Journal of Linguistics* 48 (3): 609–52. <https://doi.org/10.1017/S0022226712000011>.
- Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. "Experimental Syntax and the Variation of Island Effects in English and Italian." *Natural Language & Linguistic Theory* 34 (1): 307–44. <https://doi.org/10.1007/s11049-015-9286-8>.

- Sprouse, Jon, Carson T. Schütze, and Diogo Almeida. 2013. "A Comparison of Informal and Formal Acceptability Judgments Using a Random Sample from Linguistic Inquiry 2001–2010." *Lingua* 134 (September): 219–48. <https://doi.org/10.1016/j.lingua.2013.07.002>.
- Su, Jui-Ju, Nicola Molinaro, Margaret Gillon-Dowens, Pei-Shu Tsai, Denise H. Wu, and Manuel Carreiras. 2016. "When "He" Can Also Be "She": An ERP Study of Reflexive Pronoun Resolution in Written Mandarin Chinese." *Frontiers in Psychology* 7 (February). <https://doi.org/10.3389/fpsyg.2016.00151>.
- Tang, Chih-Chen Jane. 1989. "Chinese Reflexives." *Natural Language and Linguistic Theory* 7 (1): 93–121. <https://doi.org/10.1007/BF00141348>.
- To, Wai-Ming, and Linda S.L. Lai. 2015. "Crowdsourcing in China: Opportunities and Concerns." *IT Professional* 17 (3): 53–59. <https://doi.org/10.1109/MITP.2015.47>.
- Xu, Liejiong. 1993. "The Long-Distance Binding of Ziji." *Journal of Chinese Linguistics* 21 (1): 123–42. <http://www.jstor.org/stable/23756128>.
- Xu, Yuhang, and Jeffrey T. Runner. 2019. "Memory Retrieval in Reflexives Processing: Evidence Against the Local Search Hypothesis." *Working Papers in the Language Sciences at the University of Rochester* vol. 7 (1).
- Yu, William Xian-fu. 1992. "Challenging Chinese Reflexive Data." *The Linguistic Review* 9 (3). <https://doi.org/10.1515/tlir.1992.9.3.285>.
- . 2000. *Chinese Reflexives*. Vol. 15. Orbis. Supplementa 15. Leuven: Peeters.
- Zhan, Meilin, Roger Levy, and Andrew Kehler. 2020. "Pronoun Interpretation in Mandarin Chinese Follows Principles of Bayesian Inference." Edited by Yang Xu. *PLOS ONE* 15 (8): e0237012. <https://doi.org/10.1371/journal.pone.0237012>.