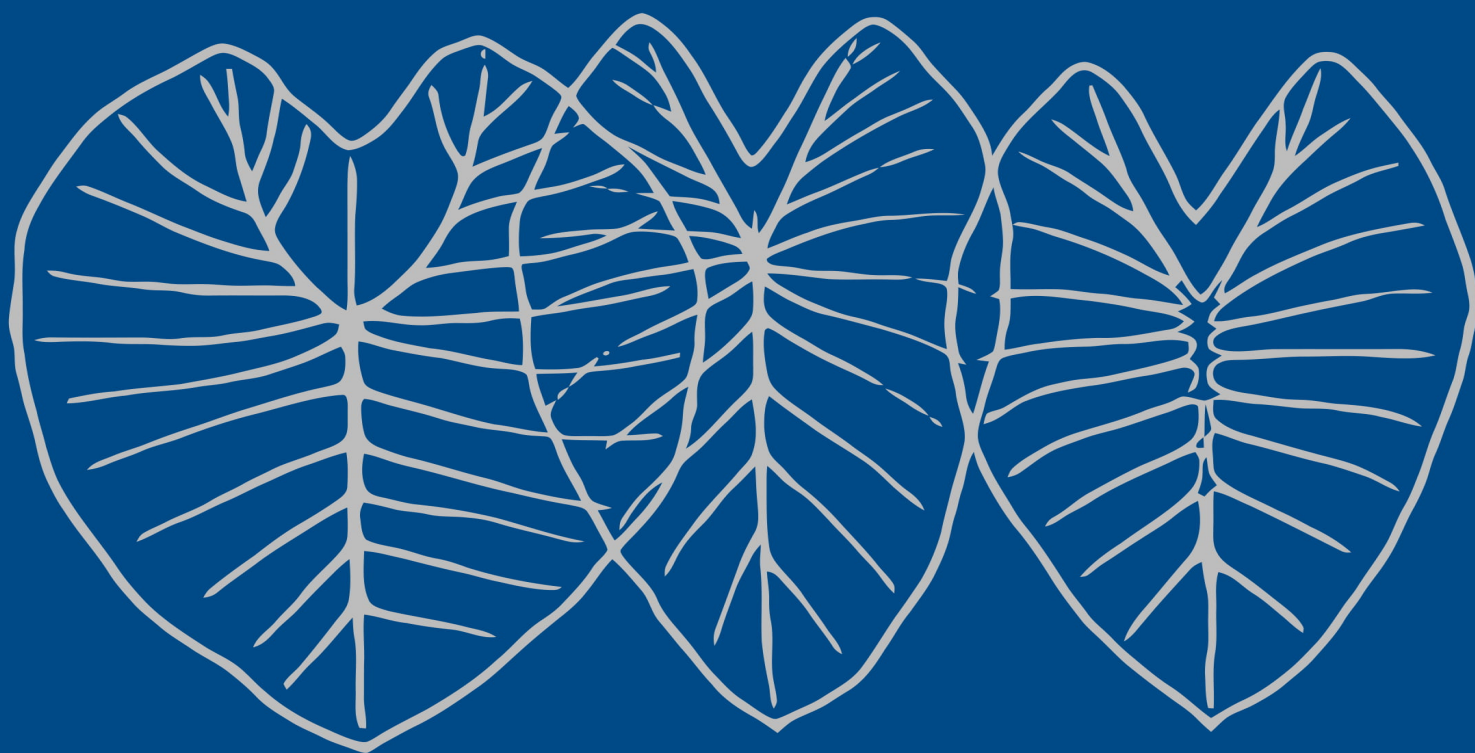# Proceedings of TripleA 4

Fieldwork Perspectives on the
Semantics of African, Asian and Austronesian Languages

University of Gothenburg
June 9-11, 2017

Edited by
Elizabeth Bogal-Allbritten and Elizabeth Coppock

Proceedings of TripleA 4:
Fieldwork Perspectives on the Semantics of African, Asian and Austronesian Languages

Edited by Elizabeth Bogal-Allbritten and Elizabeth Coppock
2018 Universitätsbibliothek Tübingen, Publikationssystem

https://publikationen.uni-tuebingen.de/xmlui/handle/10900/78334

Table of Contents

Embedded questions and concealed relative questions in Hausa and Akan[1]

Malte Zimmermann – Universität Potsdam

Abstract. The central goal of this paper is to instigate cross-linguistic research on the interpretation of embedded interrogatives and concealed relative clauses. The empirical focus is on the West African languages Hausa and Akan, which prominently employ relativized DPs for expressing embedded questions. The paper first discusses the different ways for interpreting and analyzing embedded *wh*-interrogatives: interpretations vary from strong exhaustive via intermediate and weak exhaustive to non-exhaustive. We will then present data on *concealed relative questions* in Hausa and Akan, focusing on the issues of how such structures are compositionally interpreted, and how they behave in terms of (non-) exhaustivity. Drawing on existing analyses of concealed and interrogative questions in English, we tentatively propose two formal analyses for concealed relative questions in the two languages discussed.

1   Introduction: Embedded interrogative questions in English

Embedded questions in English are proto-typically realized in the form of *yes/no*- or *wh*-interrogatives, which occur embedded under a range of predicates (Karttunen 1977):

(1)   Sigurd knew/ found out/ told us/ was surprised at **[CP who left]**.

Two central questions on embedded *wh*-interrogatives are what their semantic interpretation is eg. in terms of exhaustivity, and how this interpretation comes about. In particular, what is the meaning contribution of the embedded (interrogative) clause? What is the contribution of the embedding predicate? And what is the contribution of (covert) EXH-operators, if any? In response to these questions, the following assumptions are commonly made: (i.) Embedded questions as in (1) take the form of interrogative clauses, which denote alternative propositions (Hamblin 1973, Karttunen 1977) or partitions (Groenendijk & Stokhof 1984) at some level of their compositional semantic derivation; (ii.) The interpretation of embedded *wh*-interrogatives as (non-)exhaustive is variable, depending on context and embedding predicate. The observable surface interpretations of embedded interrogatives are derived from the interaction of alternative propositions with (lexicalized) EXH- or MAX-operators, or from the absence of such operators.

The core objective of this paper is to take a closer, cross-linguistically informed look at these basic assumptions on the meaning of embedded questions. The study of questions focuses for the most part on Western Indo-European languages (German, Romance, Slavic). But, to my knowledge, there is no systematic formal semantic work on the meaning of embedded questions in African languages, at least some of which seem to differ in the way in which embedded questions are structurally realized. For instance, Hausa (Chadic, Afro-Asiatic) has two alternative

---

ways for expressing embedded questions. First, as an embedded *wh*-interrogative (2a). Secondly, as a complex relativized DP (2b), or so-called *concealed relative question* (CRQ):

(2)   a. Musa   ya   san [wà / wàne   (ne) ya   tàfi Kano]
               Musa 3SG.M.PFV know who.SG / WHO.SG.M.   FOC 3SG.M.PFV go Kano
               'Musa knows who went to Kano.'
        b. Musa   ya   san [wa-n-dà   ya   tàfi   Kano]
               Musa 3SG.M.PFV know one.SG-DEF.M-REL. 3SG.M.PFV go   Kano
               'Musa knows who went to Kano.' (lit.'Musa knows the one that went to Kano.)

In §3, we show the CRQ-strategy to be preferred in out-of-the-blue translations into Hausa, whereas with most embedding predicates it seems to be the only available strategy for coding embedded questions in Akan (Kwa, Niger-Congo). The question is, then, how to derive embedded question meanings from the meaning of relativized DPs. More generally, the findings on Hausa and Akan point to the importance of concealed questions (CQs), as exemplified in (3). CQs are often treated as marginal in English, and yet, judging by the Hausa and Akan data, they seem to be a common cross-linguistic source for embedded question interpretations:

(3)   John knows the winners / the persons that won.   (= knows who the winners are)

Also, connecting the formal analysis of concealed questions (CQs) with embedded *wh*-interrogatives, the question arises what predictions such analyses will make for the interpretation of CRQs as strongly, weakly or non-exhaustive. This question is normally set aside in the analysis of CQs, cf. Nathan (2006:27), but will be explicitly addressed in §4.

    The paper is structured as follows. Section 2 discusses the (EXH-)interpretation and formal analysis of embedded *wh*-interrogatives in English. Section 3 introduces data from Hausa and Akan, focusing on the phenomenon of CRQs. Section 4 discusses three analyses of English CQs, focusing on the questions of how such structures are compositionally interpreted, and how they behave in terms of (non-) exhaustivity. Whereas most analyses treat CQs as strongly exhaustive, the analysis in Nathan (2006) is compatible with a weak exhaustive interpretation of CQs. This, in turn, would be compatible with variability in the EXH-interpretation of CRQs in Hausa and Akan. We will sketch a Nathan-style analysis for Hausa, juxtaposing it to a competing analysis in terms of Xiang's (2016) choice function-based analysis of *wh*-interrogatives. Section 5 concludes. Note that the results are preliminary in that data on Hausa and Akan come from only one speaker each. There is thus urgent need for further empirical confirmation; see eg. FN4.

2   Interpretation of embedded interrogatives in English: Data and Analysis

2.1   Variable Interpretation: Different degrees of (non-) exhaustivity

Embedded interrogatives allow for up to four different interpretations depending on embedding predicate and context: Strong exhaustive (SE), non-exhaustive (NE), weak exhaustive (WE), and intermediate exhaustive (IE). SE interpretations are found with the matrix predicate *know*, as in

(4a). For (4a) to be true, Sigurd knows for everybody that left that she left and he knows for everybody that didn't leave that she didn't leave. That is, Sigurd knows the complete true answer, and he knows that this is the complete true answer. Therefore, (4a) is equivalent to its negated counterpart (4b) (Groenendijk & Stokhof 1984). The SE reading is modelled in (5).

(4)   a. Sigurd knows [who left].   ⇔   b. Sigurd knows [who didn't leave].

(5)   ⟦ who left$_{SE}$ ⟧^{w} = λp$_{<s,t>}$.∃x[p(w) ∧ p = λw'. x left in w' and nobody else left in w']

At the opposite end of the EXH-scale, there are NE *mention-some* questions: These typically contain a possibility modal as in (6a) (Groenendijk & Stokhof 1984, Xiang 2016: 38), and they can be modelled as denoting sets of propositions containing at least one true answer, (6b).

(6)   a. Pekka learnt [where he can purchase a Finnish newspaper].
           (eg. Pekka found out that he can buy a Finnish newspaper at Suomen Kauppa)
        b. ⟦where P. can purchase a Finnish newspaper$_{NE}$ ⟧^{w} = P$_{<st,t>}$, such that P ⊂ {λw.
           Pekka can purchase a Finnish newspaper in x| x∈SHOP}and ∀p∈P[p(w)]

NE interrogatives show indefinite behavior in denoting restricted variables, corresponding to the behavior of the indefinite determiners in (7) (Heim 1982). The semantic effect can be modelled by means of choice functions (Reinhart 1997, Xiang 2016):

(7)   ⟦a/some man⟧= f$_{CH}$(⟦man⟧) = x, such that x∈⟦man⟧

WE embedded interrogatives as in (8a), by contrast, denote propositional sets containing the complete list of true answers (and nothing else), as in (9). Because of this, WE interrogatives are not equivalent to their negated counterpart (8b). In (8a), for instance, Jaden may be surprised at everybody who is dating Karina, but not at who is not.

(8)   a. Jaden is surprised at [who is dating Karina]. ≠
        b. Jaden is surprised at [who is not dating Karina].

(9)   ⟦who is dating Karina$_{WE}$ ⟧^{w} = λp$_{<s,t>}$.∃x[p(w) ∧ p = λw'. x is dating Karina in w']

WE readings are commonly found with emotive/cognitive verbs such as *be surprised at* or *wonder* (Berman 1991, Heim 1994:139). They are also attested with the communication verbs *tell*, *remind*, and *write down* (Karttunen 1977:11, Heim 1994:137). (10) illustrates for *tell*:

(10)   John told Mary (correctly) who passed the test, …
           (≈ John gave Mary the complete list of people who left)
        (i) …, but his list mistakenly included a few students that did not.
        (ii) …, indeed, in his eagerness, he even named a few students that did not.

Finally, Spector (2006), Klinedienst & Rothschild (2011), and Uegaki (2015) claim that embedded interrogatives can also have intermediate exhaustive readings. Predications of the

form *X P [wh]* with factive or cognition verbs (*discover*) and with non-factive communication verbs (*tell, predict*) are claimed to be true if x stands in P-relation to all true answers to *wh*, and, moreover, X stand in no P-relation to any false answer to *wh*. Accordingly, (11a) will be IE-true if Arthur told us for everybody who sang that this person sang and for everbody else, Arthur did not make any false claims that this person sang. Likewise, Cremers and Chemla (2016) show in an experimental study that speakers judge (11b) as true in an IE-scenario: All blue squares are correctly predicted, and there's no false prediction that any of the non-blue squares be blue.

(11)   a. Arthur told us [who sang].
        b. John predicted [which of the four squares were blue].

In sum, embedded interrogatives may have up to four different surface readings, as informally summarized in (12), where WE stands for the set of true answers to the question. The four readings are ordered in terms of logical strength SE>IE>WE>NE: If a question clause is true on a stronger reading to the left, then all weaker readings to the right are entailed to be true as well.

(12)   a. SE: **All p∈WE are true** & all q∉WE are false (= Only p's∈WE are true)
        b. IE: **All p∈WE are true** & and ¬∃q∉WE that is falsely claimed to be true.
        c. WE: **All p∈WE are true**
        d. NE: **There is some P⊂WE**, such that the elements of P are true

In conclusion, we list a few potential confounds for the empirical investigation of the meaning of embedded interrogatives: The domain of quantification may be uncertain (George 2011, G&S 1984:87), or there may be non-complementary background predicates (George 2011:88f.). Verbs of saying like *tell* oscillate between veridicality and non-veridical interpretations (Egré and Spector 2015). The intended granularity of the answer may differ (eg. Lahiri 2002), and there may be various conceptual covers associated with possible answers (Aloni 2001). These factors must be controlled for in experiments and fieldwork on embedded question interpretation.

2.2   Formal Modelling

The literature offers various ways for modelling the (non-)exhaustive interpretationof questions. The analyses differ in the underlying interpretation assigned to embedded interrogatives as well as in the origin of the EXH-effect, if present. We can distinguish between Q-operator, lexical, and covert MAX/EXH/ANS-operator accounts, respectively.

In Q-operator approaches, the interrogative semantics is essentially determined by the workings of a left-peripheral question operator, predicting a uniform semantic interpretation. For Karttunen (1977), *wh*-interrogatives come with underlying WE-readings: The set of true propositions ($<st,t>$) is derived from the interaction of Q-operator and existential *wh*-expression:

(13)   a. ⟦wh_i Q⟧= $\lambda q_{<st>}.\lambda w_{<s>}.\lambda p_{<st>}.$ $\exists x [p(w) \wedge p = [\lambda x_i.q](x)]$
        b. ⟦who Q is dating K. ⟧ (w) = $\lambda p_{<s,t>}.$ $\exists x [p(w) \wedge p = \lambda w'.$ x is dating K. in w']

For Groenendijk & Stokhof (1982, 1984), by contrast, interrogatives come with a basic SE-interpretation: They denote a partition that is derived from the meaning of Q plus λ-abstraction over the *wh*-index. The extensional type of embedded interrogative is thus <st>.

(14)   a. $[\![wh_i \ Q]\!] = \lambda q_{<st>}.\lambda w.\lambda w'. [\lambda x_i.q(w') = \lambda x_i.q(w)]$
        b. $[\![\text{who is dating K.}]\!](w) = \lambda w'. [\lambda x.x \text{ is dating K. in } w' = \lambda x.x \text{ is dating K. in } w]$

In lexical approaches, the semantics of embedded interrogatives is largely determined by the embedding predicate. Such approaches typically assume a Karttunen-style WE-interpretation for the embedded interrogative. For Heim (1994:133), the SE-interpretation typically observed with interrogatives embedded under *know* (3) comes from the lexical meaning of this predicate: *To know a question* means to know the SE-answer to this question in the sense of G&S (1984):

(15)   a. $[\![know]\!](w)(q)(x) = 1$ iff x believes $\lambda w'[q(w') = q(w)]$ in w
        b. $[\![\text{Sigurd knows who is dating Mary}]\!](w) = 1$ iff
                Sigurd believes the proposition that the set of true answers to question intension q is the set of true answers to q in evaluation world w, i.e. he believes the complete answer and he also believes that this is the complete answer.

Crucially, *know* differs from other embedding predicates such as *be surprised at* or *tell*, which do not require access to a full SE- partition, but simply express an attitude to the conjunction of all true answers (WE), some true answer (NE), or even to the question intension q with *wonder* or *ask*; cf. (23) to (25) below. The lexical account is thus well-suited for capturing variability in the interpretation of embedded interrogatives. In the same vein, Spector (2006) proposes that embedding predicates are potentially three-ways ambiguous between SE-, WE- and IE-reading.

Other accounts model the variable EXH-interpretation of embedded interrogatives by adding covert semantic operators to a basic WE interrogative interpretation. For instance, Rullmann & Beck (1999) derive the SE-reading from the underlying WE-reading by inserting a covert maximality (MAX)-operator at LF. The function of MAX is to pick the maximal true answer from the set of true answers. Klinedienst & Rothschild (2011), assume three semantic readings for embedded interrogatives: Their basic interpretation is WE. IE- and SE-readings are derived by inserting a covert EXH-operator in matrix or embedded clause, respectively:

(16) a. **WE:** Pred [wh-interrogative]   b. **SE:** Pred [EXH [wh-interrogative]]
                                                  c. **IE:** EXH [Pred [wh-interrogative]]

The matrix subject is required to have a P-attitude towards the complete list of true answers in WE (16a), to have a P-attitude towards the exhaustified list of true answers in SE (17b)(eg. to know the full list and knowing that it is the full list), and to have an exhaustified P-attitude to the full list of true answers in IE (16c) (i.e. no P-attitude to false answers). Uegaki (2015) presents a slightly modified picture, according to which interrogatives have but two distinct LFs, namely the WE- and IE-configurations in (16a) and (16c), respectively, whereas the SE-reading is pragmatically derived. Finally, Theiler (2014) deviates from most operator accounts in assuming basic NE reading for embedded interrogatives. The WE-reading is derived by adding a [+completive] ANS(wer)-operator, the strong reading by combining ANS- and EXH-operator.

In the past 20 years or so, then, most researchers take the WE Karttunen- interpretation to be the basic reading of embedded questions, which is presumably coded in a question-mood operator. Furthermore, there is a general consensus that the [+/-EXH]-interpretation of embedded questions is to some extent flexible. Importantly, the study of embedded questions focuses largely on embedded *wh*-interrogatives in European languages. From a cross-linguistic perspective, this raises the question of whether CRQs in Hausa and Akan are as flexible as English interrogatives, or whether they have stronger underlying EXH-interpretations. This question is highly relevant because it is not clear that interrogatives are indeed the preferred strategy for realizing embedded questions across languages: as will be shown below, the preferred strategy in Hausa and Akan are CRQs. If so, the analysis of EXH-effects in embedded questions will have to rely on the analysis of such CQRs and also of free relative-based questions in English (Caponigro 2003). Differences in the interpretation of interrogatives and CRQs might then be responsible for the co-existence of two embedded question strategies in Hausa, cf. (2).

3   Concealed Relative Questions in Hausa and Akan

As already mentioned in §1, the two typologically unrelated West African languages Hausa and Akan have a second strategy for expressing question embedding, next to *wh*-interrogative clauses: CRQs in Hausa were illustrated in (2), repeated here.[2]

(2)   a. Musa   ya   san   [wà / wàne   (ne)   ya   tàfi Kano]
            Musa   3SG.M.PFV   know   who.SG / WHO.SG.M.   FOC   3SG.M.PFV   go   Kano
            'Musa knows who went to Kano.'
        b. Musa   ya   san   [wa-n-dà   ya   tàfi   Kano]
            Musa   3SG.M.PFV   know   one.SG-DEF.M-REL.   3SG.M.PFV   go   Kano
            'Musa knows who went to Kano.' (lit.'Musa knows the one that went to Kano.)

Newman (2000: 502, my emphasis) points out that "One can express the semantic equivalent *(more or less)* of an indirect question by means of a relative clause construction headed by a noun or a relative pronoun." Indeed, CRQs were offered as spontaneous translations of the corresponding *wh*-interrogatives in English, suggesting that they may indeed be the preferred choice for expressing embedded questions in Hausa.

Next to the verb *sani* 'know' in (2), CRQs were solicited as translations for *wh*-interrogatives with different verbs of cognition and communication, as illustrated in (17a) to (19a). The (b/c)-sentences show the corresponding *wh*-interrogatives to be licit as well, except under the verb *annabta* 'predict' in (19b) (H. Abubakari, p.c.):

---

[2] Hausa examples are presented in Boko script with L tone diacritics (`). Akan examples are given without tones.

(17) **Translated:** 'It *surprised* Musa [who went to Kano].'
  a. Yaa   baa Musa màmakì (game) dà [wa-n-dà   ya   tàfi Kano]
   3SG.M.PFV give Musa surprise about P one.SG-DEF.M-REL 3SG.M.PFV go Kano
  b. Musa ya   **yi màmakì-n** [**wàne (nè) ya/wàce(cè)** **ta**   tàfi Kano]
   Musa 3SG.M.PFV do surprise-of who.M.SG 3SG.M.PFV/who.F.SG 3SG.F.PFV go Kano
  c. Musa ya   **yi màmakì-n** [**wà** ya/ta   tàfi Kano]
   Musa 3SG.M.PFV do surprise-of who.SG 3SG.PFV.M/F go Kano

(18) **Translated:** 'Musa *told* us [who went to Kano].'
  a. Musa ya   fàd'a manà [**wa-n-dà**   ya   tàfi Kano]
   Musa 3SG.M.PFV tell 1PL-IO one.SG-DEF.M-REL 3SG.M.PFV go Kano
  b. Musa ya   fàd'a manà [**wàne (nè)** ya/**wàce(cè)** ta   tàfi Kano]
   Musa 3SG.M.PFV tell 1PL-IO who.M.SG 3SG.M.PFV/who.F.SG 3SG.F.PFV go Kano
  c. Musa ya   fàd'a manà [**wà** ya/ta   tàfi Kano]
   Musa 3SG.M.PFV tell 1PL-IO who.SG 3SG.PFV.M./F. go Kano

(19) **Translated:** 'Musa *predicted* [who went to Kano].'
  a. Musa ya   annabtà gàme dà [**wa-n-dà**   ya   tàfi Kano]
   Musa 3SG.M.PFV predict about one.SG-DEF.M-REL 3SG.M.PFV go Kano
  b.*?Musa ya   annabtà (gàme dà) [**wà** ya/ta   tàfi Kano]
   Musa 3SG.M.PFV predict about who.SG 3SG.PFV.M./F. go Kano
  Intended: 'Musa predicted who went to Kano'

The infelicity of (19b) is surprising from the perspective of English CQs, for which the following generalization holds (Aloni & Roelofsen 2011: 471): "A generalization that emerges […] is that any verb that selects for CQs also selects for *wh*-complements"; cf. also Baker (1968). By contrast, the Hausa verb *annabta* selects for a relative concealed question, but not for a *wh*-complement. The variable realization of embedded questions as *wh*-interrogatives or CRQs in Hausa raises a more general question about the interpretive potential of the two structures: Are they equivalent in expressive power? Or do they differ in terms of answerhood conditions, presuppositions, or exhaustivity? Empirically, semantic equivalence would predict free substitutability *salva veritate*, i.e. a parallel distribution. Differences in semantic interpretation would be correlated with differences in syntactic and contextual distribution.

  Notice that the difference between *wh*-interrogatives and CRQs does not per se lie in the fact that the latter are overtly specified for [number] and [gender] features, presumably in the form of a presupposition restricting question domain and answer space. While this additional information certainly adds to the expressive force of CRQs, we can observe that the same is possible with the *wh*-interrogatives in (2) and (17b/18b), in which number and gender of the individual(s) in question are explicitly specified by *wane ne* (M.SG), *wace ce* (F.SG) and *suwa(ne ne)* (PL) and by the relative form of the preverbal person-aspect marker (Newman 2000).[3] If anything, Hausa displays a split between number-gender specified *wh*-interrogatives and their underspecified counterparts (with bare *wa* 'who') in the (c)-clauses: The gain in expressive power by

---

[3] It is possible that the gender-number specification expressed by *nee/cee* is responsible for the emergence of exhaustivity effects with *wh*-interrogatives and focus constructions with *nee/cee* (Hartmann & Zimmermann 2007).

number/gender-specification is thus not restricted to CRQs, so that these cannot be preferred on grounds of greater informativity. A similar situation obtains in English, which also displays a split between obligatorily specified CQs (20a) and *which*-NP-interrogatives (20b), on the one hand, and unspecified bare *wh*-interrogatives, on the other (20c) (Xiang 2016).

(20)   a. John found out [the winner/s].   specified for [SG] / [PL]
         b. John found out [which participant/s won].   specified for [SG] / [PL]
         c. John found out [who won].   unspecified for [number]

The observable parallels between Hausa CRQs and the English concealed question in (20a) raise the question of whether CRQs in Hausa (and Akan) come with the specific characteristics of concealed DP-questions in English, to be discussed in §4. This position receives intuitive support from the observation that the presence of restrictive relative modifiers supports the semantic construal of embedded DPs as CQs in English, too (Nathan (2006:118f.).

    CRQs are also regularly found in Akan, cf. (21a). Embedded *wh*-interrogatives are possible as well, at least with the embedding predicate *ask* (Saah 1994:77f.), (21b):

(21)   a. Kwadwo **nim** [**nipa** **ko** [**aa** ɔ-kɔ-ɔ   Kumase]].
         Kwadwo know person one   REL 3SG-go-PAST Kumasi
         'Kwado knows the person that went to Kumasi.' (≈ K. knows who went to K.)
        b. Mary **bisa-a** [sɛ **hena** **na** o-huu   Kofi]
         Mary ask-PST COMP who FOC 3SG-saw   Kofi
         'Mary asked who saw Kofi.'

Still, the distribution of embedded *wh*-interrogatives in Akan seems much more restricted than in Hausa. Preliminary investigations suggest that CRQs under verbs other than *wonder* or *ask* have no *wh*-interrogative counterparts. Korsah (p.c.) remarks for question embedding under *know* and *tell* that "[a]s far as I can tell, Akan has only the relative strategy". (22) shows a CRQ with *tell*:[4]

(22)   Kwadwo **ka-a** [**nipa** **ko** [**aa** ɔ-kɔ-ɔ   Kumase]] kyerɛ-ɛ   Adwoa.
       Kwadwo tell-PST person one REL 3SG-GO-PST Kumasi show-PST Adwoa
       'Kwadwo told Adwoa the person that went to Kumasi.' (… who went to K.)

Same as in Hausa, the head noun of the relative clause in (21a) and (22) is specified for number (by *ko*), thereby making CRQs more informative than their *wh*-interogative counterpart in (21b). This might, at least in part, account for the general choice of CRQs over bare *wh*-interrogatives.

    The possibility of embedded *wh*-interrogatives under matrix verb *bisa* 'ask' in (21b) may be due to selectional restriction. For instance, Aloni & Roelofsen (2011:474) analyze the question-

---

[4] There appears to be some inter-speaker variation in the expression of embedded questions in Akan. A second speaker consistently rejected embedded *wh*-interrogatives as well, but he volunteered a different relative question construction involving a free relative clause without overt head NP, cf. the variant of (21a) in (i).
(i)   Kwadwo nim   [deɛ/nea ɔ-kɔ-ɔ   Kumase]
        Kwadwo know   REL_FREE 3SG-go-PAST Kumasi

embedding predicate *wonder* in (24a) as selecting for propositional concepts of type <s,<st>>. From this, they derive the incompatibility of *wonder* with CQs (23b).

(23)   Alex is wondering   a. [CP how much the milk is]   (no answer expected)
                              b. *[DP the price of milk]

The verbs *ask* and *wonder* indeed seem to express an epistemic subject's relation to an abstract question meaning, and not to the (more or less) exhaustive answer to this question. This is different with the variant *ask for* (German *erfragen*), which expresses a relation between the subject and the answer to the question *What is the price of milk*?

(24)   a. *Alex **asked** [the price of milk].[5]   b. Alex **asked for** [the price of milk].

If Akan *bisa* 'ask' turns out to be also incompatible with CRQs, we can postulate the following lexical entry for proper question-embedding predicates (*q* = intensional question meaning).

(25)   [[wonder / ask / bisa]]
        = $\lambda q_{<sst>}. \lambda x_{<e>}.\lambda w. \forall w' \in BOUL_{x,w}$: x finds out the w-true answer to q in w'.

Interestingly, the CRQs in (21a) and (22) feature the (specific) numeral marker *ko* on the head NP, and not the DEF-marker *no*. This raises the question of whether CRQs can be introduced by both indefinite and definite NPs, as shown in (26ab), with corresponding NE or WE/SE-reading:

(26)   a. with *ko*: Kwadwo knows **a** person that went to Kumasi.   (NE)
        b. with *no*: Kwadwo knows **the** person that went to Kumasi.   (WE/SE)

Alternatively, *ko* might denote a choice function,[6] which, depending on whether or not the complex relative NP denotes a singleton set, will result in NE or WE/SE-interpretation. Choice functions play a crucial role in the question analysis of Xiang (2016), to which we turn in §4.3.

4   Concealed Questions: Meaning Composition and Exhaustivity

In this section, we first introduce the semantic characteristics of English CQs (§4.1). As will emerge, CQs are semantically more restricted than their *wh*-interrogative counterparts. In §4.2, we briefly present three formal analyses of CQs focusing on the compositional procedure and predictions on question exhaustivity, if any. As will be shown, any analysis of CQs will have to resort to some sort of type-shift in order to raise the basic DP denotation to the propositional level. Finally, we will sketch two possible analyses for Hausa and Akan CRQs in §4.3, one based on the CQ-analysis of Nathan (2006), the other on the *wh*-interrogative analysis in Xiang (2016).

---

[5] But see Nathan (2006:44, ex.29g) for a diverging grammaticality judgment.
[6] See e.g. Renans (2017) for a choice function-analysis of indefinite numeral *ko(me)* 'one' in related Gaa (Kwa).

4.1 Semantic characteristics of English concealed questions

In English, many attitude verbs allow for CQs in combination with object DPs. For instance, the prominent interpretation of (27a) is that Marc knows what the price is. And (27b) has a reading on which Marc knows which price Fred knows (without necessarily knowing the amount). In general, CQs can denote into the same semantic dimensions as *wh*-interrogatives, cf. (28):

(27) a. Marc knows [_DP the price]   NOT: *'Marc knows 3 kronor.'
      b. Marc knows [_DP the price that Fred knows].

(28) a. John knows the president of the US.   = who the POTUS is
      b. John knows the meeting place.   = where the meeting place is
                  …
      c. John knows the manner in which to succeed   = how to succeed

As already discussed in connection with Hausa (19), English CQs can be embedded under a subset of predicates selecting for wh-interrogatives (Nathan 2006:44, Roelofsen & Aloni 2011:471), including *know*, *forget*, *learn*, *discover*, *tell*, *show*, *decide*, *predict*, *investigate* etc.,
    Importantly, (definite) CQs in English always denote specificational or identity questions (Nathan 2006, Romero 2007): CQs are questions about the identity of the semantic value of a definite description, i.e. an individual concept (29, 30a). CQs never denote predicational questions about properties of contextually specified individuals, such as eg. *What is x?*, (30b).

(29) ⟦ CQ ⟧ =   ? ⟦ the NP⟧ = $x_{\langle s,e\rangle}$

(30) a. Margret knows [the capital of Italy]. = Margret knows that the coI is Rome.
         (QUD: The capital of Italy is what city?)
      b. Margret knows [Rome].   (no CQ interpretation)
         ≠ Margret knows that Rome is the capital of Italy. (QUD: What is Rome?)

Nathan (2006:21) concludes that "insofar as a concealed question denotes a question, that question is an identity question, i.e. one of the form who X is or what X is".[7] Because of this semantic restriction to specificational question meanings, English CQs have a narrower distribution than *wh*-interrogatives. Moreover, English CQs do not allow for deictic reference even on their specificational readings, unlike *wh*-interrogatives. Nathan (2006:22) observes that *which one DP is* is not a possible meaning of CQs, accounting for the infelicity of (31):

(31) I bought milk at the store (with a couple of other things). The receipt lists three
      unnamed items: one cost $1.49, the second cost $1.99, and the third cost $2.49.
      #I don't know [the price of milk]. (cf. I don't know which one the price of milk is.)

---

[7] CQs pattern with *it*-clefts, which allow for specificational, but not for predicational interpretations (Percus 1997).

Finally, CQs are semantically more restricted than identificational *wh*-interrogatives in not allowing for *de re*-interpretations (Greenberg 1977). (32b) must be interpreted de dicto.

(32)   a. John found out [who the murderer of Smith was].
        eg. 1 if John found out the identity of The Strangler (who also murdered Smith)
      b. John found out the murderer of Smith.

To sum up, English CQs are more restricted in semantic interpretation and distribution (under embedding predicates) than their wh-interrogative counterparts. This directly accounts for the limited occurrence of CQs and the preference for *wh*-interrogatives in English. Moreover, it would account for why *wh*-interrogatives exist side by side with CQRs in Hausa: If the latter are more restricted semantically, the expression of other interpretations might rely on *wh*-interrogatives. Still, this does not account for why CRQs are preferred out of the blue in Hausa, unless embedded questions are identificational questions per default. Also, it does not account for why relative questions are the only available option in Akan in most cases. We need to establish whether the restriction to CRQs in Akan entails a corresponding loss in expressivity.

4.2   Compositional derivation and exhaustivity of English CQs

This section gives a brief overview of three recent formal analyses of CQs in English, focusing on the compositional derivation and the predicted level of exhaustivity. The analyses in Romero (2007) and Aloni & Roelofsen (2011) appear to predict CQs to come with obligatory SE-readings (at least under the verb *know*). This is because they resort to the workings of maximality or partitioning (SE-) operators in deriving propositional question meanings from underlying individual concepts denoted by DPs. The analysis in Nathan (2006), by contrast, appears more flexible re exhaustivity, as type-shifting applies to (relative) NP-meanings, and the maximality operator denoted by the definite article only applies at the propositional level.

    For Romero (2007), English CQs denote proposition intensions ($<$s,st$>$). The CQ-meaning is derived by letting the partitioning operator ANS_{STR} of type $<$se,$<$s,st$>>$ in (34a) apply to the individual concept meaning of the (definite) DP, thereby raising it to propositional level. Because of partitioning, the resulting CQ comes with an SE-interpretation (34bc).

(33)   [The price of milk] is known to John.

(34)   a. $ANS_{STR} = \lambda y_{<s,e>}\lambda w\lambda w'.$ **$y(w') = y(w)$**   ($y = w^*.\iota xe[price(x,milk,w^*)]$)
        b. ⟦ [ANS_{STR} The price of milk]$_{CQ}$ ⟧
            $= \lambda w\lambda w'.$**$[\lambda w^*.\iota xe[price(x,milk,w^*)](w') = \lambda w^*.\iota xe[price(x,milk,w^*)](w)]$**
            $= \lambda w\lambda w'.$**$[\iota xe[price(x,milk,w')] = \iota xe[price(x,milk,w)]$**]
        c. ⟦ [ANS_{STR} The price of milk]$_{CQ}$ is known to John ⟧
            $= \lambda w.\forall w' \in Dox_j(w)$ **$[\iota xe[price(x,milk,w')] = \iota xe[price(x,milk,w)]]$**

Romero's proposal leaves open the possibility that DP$_{CQ}$-meanings also combine with an ANS_{WE}-operator, but this would still require some type-shifting of the individual concept DP-meaning to propositional level. Assuming the operator meaning in (35), (33) would come out true iff John knows that there is a unique maximal price of milk, and what this price of milk is.

(35) $ANS_{WE} = \lambda y_{<s,e>}\lambda w.\lambda p_{<st>}. \exists x[\, p(w) \wedge p = \lambda w'.\, x = y(w')]$ [cf. Nathan 2006:81]

It follows that John knows that nothing else is this maximal price of milk: The SE-effect persists in (35) because the individual concept meaning of DP_{CQ} interacts with semantic identification.[8]

The cover-based pragmatic analysis of CQs in Aloni & Roelofsen (2011:451-2) resembles Romero's analysis in important ways. SE is directly built into the meaning of a partitioning operator '?', cf. (36a), which is present in all CQs. And again, a type-shifting operator, namely '↑' in (36b) applies to the individual concept meaning of the DP, which is subject to pragmatic cover resolution. The operator $?x$ in (36a) picks out all and only those worlds v in which φ is assigned the same truth value as in w under the same cover resolution of x to c. If P in (36b) is contextually resolved to the identity relation $\lambda y.\, y = z$, and z to the (default) naming cover, this gives rise to the identity question reading. On this pragmatic resolution, (37a) would come out as true, for instance, iff John knows (K_j) which element from the naming cover *[Rome, Athens, ...]* is identical to the individual concept *ɩx.x is capital of Italy*, cf. (37b), The interaction of *?*-operator, identification and DP-meaning ($<s,e>$) yields an SE-reading.

(36)   a. $[\![\, ?x.\varphi \,]\!]_{M,w,gR} = \{v \mid \forall c \in R(x): [\![\, \varphi \,]\!]_{M,w,gR[x/c]} = [\![\, \varphi \,]\!]_{M,v,gR[x/c]}\}$
         b. $\uparrow_{(z,P)} \alpha = ?z.P_{<se,t>}(\alpha)$; P a contextually given predicate of individual concepts

(37)   a. John knows the capital of Italy.
         b. $K_j(?z.[\lambda y_{<se>}.y = z](ɩx.\, x \text{ is capital of Italy})) = K_j(?z.z = ɩx.x \text{ is capital of Italy})$

The analysis in Nathan (2006) differs regarding the semantic type of CQs and the compositional derivation. For Nathan, CQs denote unique propositions ($<st>$), derived by type-shifting the meaning of relational NPs ($<e,et>$) to propositional level ($<st,t>$). The DEF-operator applies at the propositional level, picking the unique proposition in a given context C, cf. (38):

(38) $[\![\text{the mayor of Berlin}]\!] = ɩp_{<st>}. [[\, \exists x_e.p = \lambda w_1.[\![\text{mayor of B}]\!](x)(w_1)] \wedge C(p)]$

The DP_{CQ} in (38) denotes the unique proposition *p* such that (a) for some individual *x, p* expresses that x is A mayor of Berlin, and (b) *p* meets a contextual restriction *C*, most frequently, that it be true in evaluation world *w*. According to Nathan (2006:18), the sentence *Rajesh knows the mayor of Berlin* will then be true if Rajesh knows the proposition that Michael Müller is the mayor of Berlin. Crucially, Rajesh's knowing the unique true proposition of the form *λw.x is mayor of Berlin* does not entail his knowing that this IS the unique true proposition, at least on a *de re*-construal of (38), thereby making (38) compatible with WE-interpretations; see FN9.

As Nathan's question-type shifter applies to relational nouns only, he postulates (ibid.:19) that a DP can be a concealed question iff its head noun is relational, or, in case of non-relational head nouns, if it is modified in certain ways, eg. with a relative clause. On a type-shifted interpretation, the RC maps NP-interpretations of type $<et>$ to sets of propositions, cf. (39):

---

[8] Cf. Nathan (2006) for discussion, as well as Rullmann (1995) and Beck & Rullmann (1999) for precursor analyses with MAX-operators in *wh*-interrogatives.

(39) a. ⟦that Kim visited last month⟧ =
       $\lambda P_{<s,et>}.\lambda p_{<st>} . \exists x_e [p = \lambda w_1 . [P(w_1)(x) \wedge$ Kim visited x last month in $w_1]]$
   b. ⟦city that Kim visited last month⟧ =
       $\lambda p_{<st>} . \exists x_e [p = \lambda w_1 . [\llbracket city \rrbracket(w_1)(x) \wedge$ Kim visited x last month in $w_1]]$
   c. ⟦the city that Kim visited last month⟧ =
       $\iota p_{<st>} . \exists x_e[p = \lambda w_1.[\llbracket city \rrbracket (w_1)(x) \wedge$ Kim visited x last month in $w_1]] \wedge p(w)$

Nathan's treatment of CQs is relevant for three reasons when it comes to the analysis of CRQs in Hausa and Akan: Firstly, it offers an explicit formal analysis of the semantic role of relative clauses in CQ-formation. Secondly, the meaning of CQs does not involve partitioning (SE), nor a maximality/uniqueness condition on individual (concept)s (SE). Nathan's approach is thus compatible with a more flexible interpretation of CRQs, at least in principle, in also allowing for WE-interpretations. Finally, the analysis is not built around individual-denoting DPs, for which reason is extends easily to CQs with indefinite determiners (and NE *mention-some* interpretations). This may be useful for the analysis of CRQs containing specific indefinite determiners in Akan (21a) and (22), for instance. More generally, the different predictions on the interpretation of CQs underline the need for more empirical work on the (non-)exhaustivity of CRQs. More empirical information on the interpretation of CRQs will certainly help to decide as to whether Nathan's analysis applies to them, or not. We will sketch such an analysis next.

4.3   Suggestions for possible analyses of the CQRs in Hausa and Akan

Applying the analysis in Nathan (2006) to the syntactic structure in (40) yields the semantic derivation in (41) for the CRQ from (2b) above.[9] (42) gives the meaning of the full clause:

(40)   [$_{DP}$ wa$_1$ [$_{DP}$ `n [$_{NP}$ t$_1$ [$_{CP}$ dà [$_{TP}$ ya táfi Kano]]]]]

(41)   a. ⟦RC⟧   = ⟦dà ya táfi Kano⟧
                        = $\lambda P_{<s,et>}.\lambda p_{<st>} . \exists x_e [p = \lambda w_1.[P(w_1)(x) \wedge$ x went to K. in $w_1]]$
   b. ⟦NP⟧ = ⟦wa⟧   = $\lambda w.\lambda x.$ x is a person in w
   c. ⟦NP RC⟧   = $\lambda p_{<st>}. \exists x_e[p = \lambda w_1.[person'(w_1)(x) \wedge$ x went to K. in $w_1]]$
   d. ⟦DEF NP RC⟧   = $\iota p[\exists x[p = \lambda w_1.person'(w_1)(x) \wedge$ x went to K. in $w_1] \wedge p(w)]$

(42) $\llbracket(2b)\rrbracket^w$ = 1 iff $\forall w' \in DOX_{MUSA}(w)$:
                $[\iota p[\exists x_e[p = \lambda w_1.[person'(w_1)(x) \wedge$ x went to K. in $w_1]] \wedge p(w)]](w')$
                $\approx$ Musa knows the unique w-true proposition of the form 'x went to Kano' (WE)

Crucially, uniqueness in (42) is evaluated relative to evaluation world *w*, and thereby not part of the doxastically accessible propositional content: (42) represents the WE-interpretation of (2b).[10]

---

[9] In Hausa, the definite determiner `n/`r follows the NP, viz. *mutumî-n* 'man-DEF' (eg. Newman 2000). We postulate, that the order NP>DET is derived by Aboh(2004)-style movement of NP to SpecDP.

[10] The formal implementation raises a non-trivial issue: the requirement that the embedded proposition be true in evaluation world *w* is satisfied by binding the *world/situation* variable (Schwarz 2009) of the CQ-definite to *w*. As mentioned below (38), this deictic behavior of the DEF-operator gives rise to a *de re*-construal. Alternatively, the

Alternatively, CRQs, and in particular those with overt specific indefinites, such as Akan (21a) and (22), may receive an analysis in terms of Xiang's (2016) analysis of *wh*-interrogatives and free relative questions; see also Caponigro (2003). The analysis appears well-suited for a number of reasons: (i.) The core of the *wh*-interrogative denotes a property ($<e,st>$), not a propositional set. This would correspond to the meaning of NP+CP-constituents in CRQs. (ii.) The *wh*-element is of type $<e,st>$ and functions as a modifier on the topical property (= the background predicate), which is denoted by the rest of the interrogative. This would correspond to the role of head noun and relative clause in CRQs, which also combine by means of predicate modification. (iii.) The *wh*-core is headed by a covert answer operator ANS, selecting for the maximally informative proposition(s) in the answer space, and by a covert choice function-operator, which selects (one of) the maximally informative proposition(s). Applied to Akan CRQs, this choice function operator may be spelt out by the specific INDEF marker *ko*, given that such indefinites are often treated as choice-function denoting (Reinhart 1997).[11] Referring the reader to Xiang (2016: 14ff.), this would yield the structure in (43) for the CRQ in (22a):

(43)   [$_{DP}$ nipa$_1$ [$_{DP}$ ko   [ANS   [$_{NP<e,st>}$ t$_1$ [$_{CP<e,st>}$ aa   ɔ-kɔ-ɔ   Kumase]]
           person   CHOICE   REL   went   Kumasi

5   Conclusion

In light of the discussion, one may wonder about the cross-linguistic implications of the regular occurrence of CRQs in Hausa/Akan for the analysis of embedded *wh*-questions in English? The question is whether a unified analysis is possible on which English *wh*-interrogatives are reanalyzed as embedded CRQs, with the matrix predicate selecting for a free relative clause (Jacobson 1995, Caponigro 2003)? Xiang's (2016) analysis constitutes an important step towards a unified analysis, but it still leaves open a number questions: Do CRQs come with more limited interpretive options like the CQ-couterparts (see §4.1)? Do CRQs show a different exhaustivity behavior than *wh*-interrogatives? And why does the range of licit *wh*-expressions in *wh*-interrogatives and free relatives differ in English (Caponigro 2003)? There seemed to be no such difference between English *wh*-interrogatives and Hausa/English CRQs.

Deferring the question of whether a unified analysis for *wh*-interrogatives and relative-based questions is possible to future research, we conclude more generally that there is cross-linguistic (and language-internal) variation in the formal expression of embedded complement questions, namely as *wh*-interrogative questions or CRQs (or free relatives). This raises the issue of whether there is a corresponding difference in the semantic interpretation of embedded questions. However, the interpretion of concealed questions as SE, IE, WE, or NE has been little explored so far, not to mention the interpretation of CRQs in Hausa and Akan. It was shown that existing formal analyses of concealed questions differ in their predictions regarding the exhaustivity of

---

world variable may be bound by the matrix predicate, on a *de dicto*-construal. Uniqueness would then be established relative to the doxastically accessible worlds, thereby giving rise to SE-readings. Notice that the *de re* (world) analysis clashes with the *No de re*-constraint on CQs illustrated in (31) for English. More research is required.

[11] Notice that the Hausa element *wa(a)*, which is found as *wh*-element in *wh*-interrogatives and as an unspecific person head noun in CRQs, also forms part of the specific indefinite determiners *wani* (m.), *wata* (f.), *wasu* (pl.), thereby motivating a choice function analysis for Hausa CRQs as well.

questions, thereby raising the need for substantial empirical work on CRQs in Hausa and Akan and on concealed or free relative questions in English alike.

References

Aboh, E.O. (2004). The morphosyntax of complement-head sequences: Clause structure and word order patterns in Kwa. Oxford: OUP.

Aloni, M. (2001). *Quantification under Conceptual Covers*. PhD thesis. Amsterdam.

Aloni, M. & F. Roelofsen (2011). Interpreting concealed questions. *Linguistics and Philosophy* 34, 443–478. DOI 10.1007/s10988-011-9102-9

Beck, S. and H. Rullman (1999). A Flexible Approach to Exhaustivity in Questions. *Natural Language Semantics* 7, 249–298.

Berman, S. (1991). *On the Semantics and Logical Form of Wh-Clauses*. PhD thesis, University of Massachusetts at Amherst.

Caponigro, I. (2003). *Free Not to Ask: On the Semantics of Free Relatives and Wh-Words Cross-linguistically*. PhD thesis, UCLA.

Cremers, A. and E. Chemla (2016) A Psycholinguistic Study of the Exhaustive Readings of Embedded Questions. *Journal of Semantics* 33, 49–85.

Egré, P. and B. Spector (2015). Embedded questions revisited: an answer, not necessarily the answer. *Synthese* 192, 1729-1784.

George, B. J. (2011). *Question Embedding and the Semantics of Answers*. PhD thesis, UCLA.

Greenberg, B. (1977). A semantic account of relative clauses with embedded question interpretations. Ms., UCLA.

Groenendijk, J. and M. Stokhof (1982). Semantics analysis of wh-complements. *Linguistics and Philosophy* 5, 175–233.

Groenendijk, J. and M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam.

Hartmann, K. & M. Zimmermann (2007). Exhaustivity Marking in Hausa: A Re-Evaluation of the Particle *nee/cee*. In E. O. Aboh et al. (eds.), *Focus Strategies in Niger-Congo and Afro-Asiatic*. Mouton de Gruyter, Berlin. 241-263.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. PhD thesis, UMass.

Heim, I. (1994). Interrogative semantics and Karttunen's semantics for know. In R. Buchalla and A. Mittwoch (eds.), *Proceedings of IATL* 9, 128–144. Jerusalem: Academon.

Karttunen, L. (1977). Syntax and Semantics of Questions. *Linguistics and Philosophy* 1, 3–44.

Klinedinst, N. and D. Rothschild (2011). Exhaustivity in Questions with Non-factives. *Semantics and Pragmatics* 4(2), 1–23.

Lahiri, U. (2002). *Questions and Answers in Embedded Contexts*. Oxford: OUP.

Matthewson L. (2004). On the Methodology of Semantic Fieldwork. *International Journal of American Linguistics* 70(4), 369–415.

Nathan, L. (2006). *On the interpretation of concealed questions*. PhD thesis, MIT.

Newman, P. (2000). *The Hausa Language*. New Haven: Yale University Press.

Reinhart, T. (1997). Quantifier scope: how labor is divided between QR and choice functions. *Linguistics and Philosophy* 20, 335–397.

Renans, A. (2017). Two Types of Choice-Functional Indefinites: Evidence from Ga (Kwa). *Topoi*. DOI 10.1007/s11245-017-9479-3.

Romero, M. (2007). On Concealed Questions. In M. Gibson and J. Howell (Eds.) Proceedings of Semantics and Linguistic Theory XVI. pp. 208-227 Ithaca, NY: CLC.

Rullmann, H. (1995). Maximality in the Semantics of Wh-Constructions. PhD thesis, UMass.

Saah, K. (1994). *Studies in Akan Syntax, Acquisition and Sentence Processing*. PhD thesis, University of Ottawa.

Schwarz, F. (2009). Two Types of Definites in Natural Language. PhD thesis, UMass.

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques*. PhD thesis, Université Paris Diderot, Paris.

Theiler, N. (2014). *A Multitude of Answers: Embedded Questions in Typed Inquisitive Semantics*. MSc thesis, ILLC, Universiteit van Amsterdam.

Uegaki, W. (2015). *Interpreting questions under attitudes*. PhD thesis, MIT.

Focus strategies in Limbum[1]

Imke Driemel — Leipzig University
Jude Nformi Awasom — Leipzig University

Abstract. Limbum exhibits two morphologically marked focus strategies involving the particles *á* and *bá*. We show that the two focus markers differ in their functional complexity. While *á* introduces an existence presupposition operating on focus alternatives, *bá* additionally encodes an exhaustivity presupposition as well as a mirative component. The latter makes it possible for Limbum to show the mirror image of what is traditionally observed as a markedness distinction between information focus and contrastive focus: focus *á* marking is accompanied by syntactic fronting while *bá* marked constituents can be left in-situ. Limbum, furthermore, shows that focus by itself does not presuppose existence, since this restriction is only present when accompanied by the particle *á*, which must be left out if the context does not satisfy this presupposition.

1 Introduction

Limbum – a Grassfields Bantu language, spoken in Cameroon – shows a very extensive focus marking system, involving particles as well as syntactic re-ordering. Its basic word order is SVO, with TAM markers preceding the verb. Like most African languages, Limbum is a tone language and has three level tones (High, Mid, Low) and five contour tones (High-Mid, High-Low, Mid-Low, Low-Mid, Low-Low) which are contrastive and are marked on syllables.

(1) ŋwὲ fɔ̄   àm   tíh ŋgū
     man DET PST3 cut wood
     'The man cut the wood.'

Focus signals the presence of alternatives (Rooth 1992, Krifka 2008). Typical contexts for focus are given in (2)-(4), where (2) triggers *information focus*, while (3) and (4) are instances of *contrastive focus*. The latter is standardly distinguished from the former by adding semantic and/or pragmatic conditions on the alternatives, be it exhaustivity (Szabolcsi 1981, Kiss 1998, Vallduví and Vilkuna 1998, Horvath 2010, 2013), exclusivity (Beaver and Clark 2008, van der Wal 2011, Orenstein and Greenberg 2013, van der Wal 2014), or unexpectedness (Zimmermann 2008, Hartmann 2008, Skopeteas and Fanselow 2009, 2011, Frey 2010).

(2) A: Who stole the cookie?
     B: [PEter]$_F$ stole the cookie.   *Q-A congruence*

(3) A: Mary stole the cookie.
     B: (No,) [PEter]$_F$ stole the cookie.   *correction*

(4) An [AMErican]$_F$ farmer talked to a [CaNAdian]$_F$ farmer.   *contrast*

While information focus is often encoded by means of a canonical focus structure, the contrastive

---

[1]We would like to thank Mira Grubic and the audience of TripleA 4 in Gothenburg for helpful comments.

focus is realized with a relatively more marked focus strategy. Depending on the language, this difference can manifest itself in the opposition of in-situ vs. ex-situ structures (Hartmann and Zimmermann 2014, Fiedler et al. 2010), different levels of prosodic prominence (Bolinger 1961, Alter et al. 2001, Katz and Selkirk 2009), or the opposition of prosodic prominence and reordering/clefting (Skopeteas and Fanselow 2009). From a pragmatic point of view, the different levels of complexity receive an explanation by the observation that contrastive information often comes with an unexpectedness flavour, which, thus, requires the more marked focus strategy in order to facilitate common ground update (Skopeteas and Fanselow 2009, Zimmermann 2008, 2011, Zimmermann and Onea 2011).

Focus by itself can be left unmarked in Limbum, see (5). Note that wh-words behave completely parallel to focused constituents, supporting the hypothesis that they are intrinsically focused (Beck 2006, Haida 2007).

(5) A: wɛ̀   bí   kɔ̄nī   **ndā̄**
         you.SG FUT1 meet **who**
         'Who will you meet?'
     B: mɛ̀ bí   kɔ̄nī **Ngàlá**
         I   FUT1 meet **Ngala**
         'I will meet NGALA.'

The two marked strategies are shown in (6) and (7). The *á* strategy seems similar to cleft constructions, where the focus marker appears clause initially, followed by an optional complementizer,[2] and the focused constituent. In contrast, the particle *bá* co-occurs with focused constituents in-situ.

(6) A: á   **ndá** wɛ̀   bí   kɔ̄nī
         FOC **who** you.SG FUT1 meet
         'Who is it that you will meet?'
     B: á   **Ngàlá** (cí)   mɛ̀ bí   kɔ̄nī
         FOC **Ngala** (COMP) I   FUT1 meet
         'I will meet NGALA.'

(7) A: wɛ̀   bí   kɔ̄nī bá **ndá**
         you.SG FUT1 meet FOC **who**
         'Who (if not X) will you meet?'
     B: mɛ̀ bí   kɔ̄nī bá **Ngàlá**
         I   FUT1 meet FOC **Ngala**
         'It is Ngala whom I will meet.'

We will show that the *á* strategy is compatible with information focus, while the *bá* strategy shows signs of contrastive focus, i.e. exhaustivity and unexpectedness. The pattern, thus, instantiates the exact mirror image to the standard dichotomy, which is that contrastive focus tends to be more marked than new information focus. Pragmatic reasoning can therefore not be the source of different levels of markedness in Limbum. Hence, we propose that the notion of unexpectedness is directly encoded in the focus marker *bá*. Section 2 will demonstrate how unmarked focus differs from focus marked by the particle *á*, while section 3 develops an analysis for the particle *bá*. In section 4 we compare contexts which in principle should allow for both marked strategies but where only one of them is felicitous. Section 5 wraps up.

---

[2]Glossing *cí* as COMP is an oversimplification. See Becker et al. (to appear) for a syntactic analysis that takes *cí* as the head of a left peripheral focus projection.

2 Existence focus

Both unmarked focus and *á* marked focus are compatible with question-answer contexts, an additional example is given in (8).

(8) **Context**: Tata comes across a dead animal in the backyard. He shows it to Yaah and asks:

    Tata: (á) **ndā** à   zhvū nyà
              FOC **who** 3SG kill   animal
              'Who killed the animal?'
    Yaah: (á) **Nfò** à   zhvū nyà
              FOC **Nfor** 3SG kill   animal
              'NFOR killed the animal.'

We conclude that the presence of alternatives does not have to be marked morphologically.[3] The subtle difference between the strategies lies in the presence of an existence presupposition with the latter, but not with the former. The following context ensures that the proposition is true for at least one alternative, thus an existence presupposition is satisfied. The particle *á* is required in such a case.

(9) **Context**: Tata comes across a dead animal in the backyard. The animal appears to have been killed by someone since it shows multiple knife wounds. He shows it to Yaah and asks:

    Tata: (á)[#] **ndā** à   zhvū nyà
              FOC **who** 3SG kill   animal
              'Who is it that killed the animal?'
    Yaah: (á)[#] **Nfò** à   zhvū nyà
              FOC **Nfor** 3SG kill   animal
              'NFOR killed the animal.'

Following the work of Rooth (1985, 1992), we implement this observation in the framework of *alternative semantics*, see (10). Focus marker *á* associates with focus alternatives and introduces an existence presupposition that operates on those alternatives.[4]

(10) $[\![\text{á } \phi]\!]^o = \lambda w : \exists p[p \in [\![\phi]\!]^f \land p(w) = 1].[\![\phi]\!]^o(w) = 1$

For intonational languages such as English, the possibility of an existence presupposition is still under debate (Dryer 1996, Rooth 1999, Geurts and van der Sandt 2004, Büring 2004). The dialogue in (11) tests for presupposition status and suggests that focus on its own cannot introduce an existence presupposition. Since the context assures that Peter doesn't know if somebody saw John, an existence presupposition would not be satisfied, nevertheless intonational focus is felicitous. The control structure is a cleft which uncontroversially introduce an existence presupposition (Percus 1997, Velleman et al. 2012) and is, thus, infelicitous.

---

[3]Another context in which one would expect focus marking to occur obligatorily is under the scope of focus sensitive adverbs such as only. Again, *á* is optional, as will become apparent in the next section, see (18)A.

[4]Another way to introduce an existence presupposition is by forming the disjunction of the propositions in the alternative set: $[\![\text{á } \phi]\!]^o = \lambda w : \bigcup[\![\phi]\!]^f = 1.[\![\phi]\!]^o(w) = 1$

(11) Did anyone see John?   *(Dryer 1996: 490)*

    Peter: I don't know. I know MARY didn't see him.

    Peter': I don't know. # I know it wasn't MARY that saw him.

Again, Limbum seems to make a clear distinction in that the *á* strategy patterns like the cleft in (11). In (12), Yaah's answer is infelicitous if focus marker *á* is present because the existence presupposition contradicts the fact that Yaah answers the question if someone is playing the drums with *I don't know*. In other words, the existence presupposition is stable under negation.

(12) Shey: mè   shī   yōʔ yū   mɔ́ʔ. ŋwɛ̀   mɔ́ʔ shī   bōʔ ncùh à

           1SG PROG hear thing one   person one PROG play drum Q

           'I heard something. Is somebody playing the drums?'

    Yaah: mè   rɨ̀ŋ   kāʔ mè   shī   kwàʔshī nɛ̄ (#á) **Nfò** í   shì   bōʔ ncùh kâʔ

           1SG know NEG 1SG PROG think   that FOC **Nfor** 3SG PROG play drum NEG

           'I don't know. But I don't think NFOR is playing the drums.'

A similar projection test can be constructed with a modal operator, see (13). As above, the presence of *á* renders the answer infelicitous.

(13) Shey: ŋwɛ̀   mɔ́ʔ à   mū   lō   yá   sàʔ   á

           person one 3SG PST2 borrow 1SG.POSS cutlass Q

           'Did someone borrow my cutlass?'

    Yaah: mè   rɨ̀ŋ   kāʔ kàdéʔ bā nɛ̄ (#á) **Nfò** à   mū   lō

           1SG know NEG can   be that FOC **Nfor** 3SG PST2 borrow

           'I don't know. It is possible that NFOR borrowed it.'

Additional evidence comes from the fact that *á* cannot scope over quantifiers that denote the empty set. This is predicted since such an assertion would contradict an existence presupposition.

(14) (*á) **ŋwɛ̀**   mɛ mū   yɛ̄ káʔ

      FOC **person** 1SG PST2 see NEG

      'I saw NOBODY.'

This section has shown that focus itself does not have to be morphologically marked. The particle á introduces an existence presupposition that operates on the alternatives of the focused constituent. Limbum, thus, provides a unique window into the discussion of existence focus, in that it ties the existence presupposition to an additional marker – a counter-argument to theories that take focus alone to be the reason for the presupposition (Geurts and van der Sandt 2004).

3   Exhaustive, mirative focus

A good way to illustrate how *bá* is different from *á* and unmarked focus is by comparing the context in (15) to the contexts in (8) and (9). The context in (15) introduces an unexpectedness component, both on the hearer's and on the speaker's side. The *bá* strategy is the only option here – a pattern that clearly contrasts with (8) and (9) where the *bá* strategy is excluded.

(15) **Context**: Tata comes across a dead animal in the backyard. He immediately suspects Shey to have killed the animal but it turns out that Shey is not the culprit. Shey knows that Tata suspected him although it was Nfor who killed the animal. Tata shows the animal to Shey and asks:

Tata: #(á) **ndā** à zhvū nyà
FOC **who** 3SG kill animal
'Who is it that killed the animal?'

Tata′: à zhvū bá **ndā** nyà
EXPL kill FOC **who** animal
'Who (if not you) killed the animal?'

Shey: #(á) **Nfò** à zhvū nyà
FOC **Nfor** 3SG kill animal
'NFOR killed the animal.'

Shey′: à zhvū bá **Nfò** nyà
EXPL kill FOC **Nfo** animal
'It is Nfor who killed the animal.'

The exhaustive component can be exemplified with correction scenarios and co-occurrence restrictions with additive particles such as *also* (see Becker and Nformi 2016). A correction context requires an utterance with an explicit alternative, followed by another alternative in a second utterance, automatically canceling the first one. In such contexts, the *bá* strategy is obligatory.

(16) **Context**: Ndi bought a pair of shoes. Njobe does not remember correctly and tells Tanko that Ndi bought a dress. Ndi corrects Njobe saying that she bought shoes (instead).

Njobe: í bá yū bcɛ̀ʔ
she PST2 buy dresses
'She bought dresses.'

Ndi: #mɛ̀ bá yū **blábáʔ**
I PST2 buy **shoes**
'I bought SHOES.'

Ndi″: mɛ̀ bá yū bá **blábáʔ**
I PST2 buy FOC **shoes**
'It is shoes that I bought.'

Ndi′: #á **blábáʔ** (cí) mɛ̀ bā yú
FOC **shoes** (COMP) I PST2 buy
'I bought SHOES.'

Exhaustivity also prevents *bá* from co-occurring with the focus sensitive adverb *fɔ́ŋ* 'also', since additives[5] require a proposition to be true for at least one non-selected alternative, whereas *bá*

---

[5]The scalar additive *káʔ* 'even' seems to be restricted to topics rather than foci since it cannot co-occur with *á* or *bá* and is, thus, untestable for our hypotheses. As (i) shows, *káʔ* has to precede the focused constituent, while *á* is illicit independent of where exactly in the left periphery the particle occurs. A similar pattern can be shown for *bá*, albeit with the additional restriction that constituents under the scope of *káʔ* have to undergo fronting, see (ii).

(i) (*á) káʔ (*á) **Ngàlá** (*á) mɛ̀ bí kɔ̄nī
FOC even FOC **Ngala** FOC I FUT1 meet
'I will meet even NGALA.'

(ii) *mɛ̀ bí kɔnī (*bá) káʔ (*bá) **Ngàlá** (*bá)
I FUT1 meet FOC even FOC **Ngala** FOC
'It is even Ngala whom I will meet.'

In (iii) we show that topics in general need to be fronted and can optionally leave a resumptive pronoun – both of which is true for constituents under the scope of *káʔ*, see (iv) and (v). Focused constituents marked with *á* require fronting as well. A resumptive pronoun, however, is not allowed to show up, see (vi).

requires all non-selected alternatives to be false.

(17) Shey: Nfò à   mū   yū rkār.
           Nfor 3SG PST2 buy car
           'Nfor bought a car.'

   Ndi: í   mū   yū **ntùmntùm** fɔ́ŋ.   Ndi″: *í   mū   yū bá **ntùmntùm** fɔ́ŋ.
          Nfor PST2 buy motorbike   also   3SG PST2 buy FOC motorbike   also
          'He bought a MOTORBIKE also.'   'It is a motorbike he also bought.'

   Ndi′: á   **ntùmntùm** (cí)   í   mū   yū fɔ́ŋ.
          FOC motorbike   COMP 3SG PST2 buy also
          'He bought a MOTORBIKE also.'

To show that mirativity is at work independently of exhaustivity, we show the behaviour of *bá*
under the scope of another exhaustive operator *càʔcàʔ* 'only'. If one ensures exhaustivity with
càʔcàʔ, *bá* becomes licit only if the selected alternative is also unexpected.

(18) **Context:** Shey is looking for Ngala and Tanko who are supposed to be at the market. Shey
     tells Ndi to go find Ngala and Tanko and bring them back.

(iii) mbǎ   fɔ̄, wōyè ó   ∅   fā   (zhí) nì   yē wéé
      money DET they 3PL PERF give it.RES PREP 3SG already
      'The money, they already gave it to him/her.'

(iv) káʔ **Ngàlá** mɛ̀ bí   kɔ̄nī (yē)
     even **Ngala** I   FUT1 meet 3SG.RES
     'I will meet even NGALA.'

(v) *mɛ̀ bí   kɔnī káʔ **Ngàlá**
    I   FUT1 meet even **Ngala**
    'It is even Ngala whom I will meet.'

(vi) á   **Ngàlá** mɛ̀ bí   kɔ̄nī (*yē)
     FOC **Ngala** I   FUT1 meet 3SG.RES
     'I will meet NGALA.'

We take this as evidence that *káʔ* can only scope over topics. Hence, *káʔ* is incompatible with the *á* strategy, which is
shown in (i). Since topics have to be fronted, (ii) is unacceptable, independent of the presence of *bá*.

A. Yaah comes back with Ngala.

Yaah: mɛ̀ ∅ kɔ́ní **Ngàlá** cà?cà?
I PERF find **Ngala** only
'I only found NGALA.'

Yaah′: á **Ngàlá** cà?cà? (cí) mɛ̀ ∅ kɔ́ní
FOC **Ngala** only COMP I PERF find
'I found NGALA only.'

Yaa″: #mɛ̀ ∅ kɔ́ní bá **Ngàlá** cà?cà?
I PERF find FOC **Ngala** only
'I only found NGALA.'

B. Yaah comes back with Njobe.

Yaah: mɛ̀ ∅ kɔ́ní bá **Njobe** cà?cà?
I PERF find FOC **Njobe** only
'I only found NJOBE.'

Both the exhaustive and the mirative component seem to be non at-issue, as the following two tests suggest. For the continuations in (19) to be informative (and thus felicitous), exhaustivity – encoded by either *bá* or *cà?cà?* – must be at-issue. As (19)i shows, the *bá* continuation is infelicitous, i.e. *bá* asserts the ordinary semantic value of the focused constituent, while presupposing an exhaustified focus alternative set. In contrast, (19)ii suggests that *cà?cà?* asserts exhaustivity, similar to its English counterpart *only*.

(19) mɛ̀ r̀ŋ nɛ̄ Tata à mū zhē **mŋgɔ̀mbé** kāp kū yō? nɛ̄ ...
1SG know that Tata 3SG PST2 eat **plantain** but just hear that
'I know Tata ate PLAINTAIN but I've just heard that...'

(i) *bá*: exhaustivity not at-issue

... #í mū zhē bá **mŋgɔ̀mbé**
3SG PST2 eat FOC **plantain**
'it was PLAINTAIN she ate.'

(ii) *cà?cà?*: exhaustivity at-issue

... í mū zhē **mŋgɔ̀mbé** cà?cà?
3SG PST2 eat **plantain** only
'she only ate PLAINTAIN.'

The non at-issue status of the mirative component is suggested by the fact that it can project out of the antecedent of conditionals. Compare (20) to (21), where adding *bá* in (21) lets the mirative interpretation of the selected focus alternative project through, so that it escapes cancellation in the consequent. This is not true for (20), due to the absence of *bá*.

(20) [kā? bā nɛ̄ í mū nɔ̄ **mbrò?**] mɛ̀ lɛ̄? kā?
if to.be that 3SG PST2 drink wine 1SG surprise NEG
'If it is WINE that he drank, I'm not surprised.'
Comment: The lack of surprise can be directed at the choice of beverage.

(21) [kā? bā nɛ̄ í mū nɔ̄ bá **blē̄ē**] mɛ̀ lɛ̄? kā?
if to.be that 3SG PST2 drink FOC blood 1SG surprise NEG
'If it is BLOOD that he drank, I'm not surprised.'
Comment: The lack of surprise can only be directed at the person, given that this person does unusual things all the time. It cannot be directed at blood.

Besides unexpectedness and exhaustivity, *bá* additionally encodes existence (just like *á*), shown here by the inability to occur with a negative quantifier, see (22).

(22) Tanko à   mū yū bflǎwà Ngàlá à   yu (*bà) **yū**   kāʔ
Tanko 3SG PST2 buy flowers Ngala 3SG buy FOC thing NEG
'Tanko bought flowers but Ngala bought NOTHING.'

Existence is presupposed, since it can project through negation, compare (23) to (24). Once a focused constituent is preceded by *bá*, the existence of the selected alternative cannot be denied, even if *bá* is embedded under negation.

(23) [**Tata** à   mū sɔ̄ mbàŋ káʔ] àndʒɔ́ʔ ŋwɛ̀ mɔ̀ʔ à   mū sɔ̄ kāʔ
 Tata 3SG PST2 win game NEG because person one 3SG PST2 win NEG
'TATA did not win the game because nobody won.'

(24) *[à   mū sɔ̄ bá **Tata** mbàŋ káʔ] àndʒɔ́ʔ ŋwɛ̀ mɔ̀ʔ à   mū sɔ̄ kāʔ
 EXPL PST2 win FOC Tata game NEG because person one 3SG PST2 win NEG
'It is not TATA who won the game because nobody won.'

Since we would like to model the exhaustivity as well as the existence requirement as a presupposition, we run into the problem of making the entailed content look trivial – a problem which has been discussed for English clefts (Velleman et al. 2012).

(25) It was Mary who laughed.

   a. $\exists x[laughed(x)]$ *existential presupposition*
   b. $\forall x[laughed(x) \rightarrow (x = m)]$ *exhaustive presupposition*
   c. $laughed(m)$ *entailed prejacent*

An additional problem relates to the observation that the exhaustive presupposition does not project through negation. This is true for clefts (Velleman et al. 2012, Büring and Križ 2013), see (26), as well as for the *bá* strategy, shown in (27).

(26) It wasn't Mary who laughed; it was Bill.

(27) mè bí   kɔ̄nī bá **Ngàlá** ká
1SG FUT1 meet FOC **Ngala** NEG
'It is not NGALA I will meet.'

... mè bí   kɔ̄nī bá **Nfò**
    1SG FUT1 meet FOC **Nfor**
    'It is NFOR that I will meet.'

We would like to follow Büring and Križ (2013) who offer a solution towards these problems by making the exhaustive presupposition dependent on the assertion. Thus, (25-b) has to be reformulated along the lines of *If Mary laughed, then nobody else did.* Büring and Križ (2013) make use of a *max* operator in their exhaustivity presupposition, which is based on mereological parthood, see (28) and (29). For the cleft in (30) for example, the *max* operator derives the sum of all invitees, given that all predicates are always closed under fusion and therefore each have a maximal element.

(28) $max(P) = \{\, x \in P \mid \neg \exists y \in P[x \sqsubset y] \,\}$

(29) CLEFT:= $\lambda z.\lambda P : \forall x \in max(P)\ [z \not\sqsubseteq x].P(z)$

An example for a positive cleft is given in (30). Following the denotation of (29), the presupposition, given in (30-b), has two inferences: (i) Fred is the sole invitee, (ii) Fred is not invited. Given the assertion in (30-a), only (ii) is a licit inference. Hence, an exhaustive effect arises.

(30) It was Fred she invited.

    a.   ASS: She invited Fred.

    b.   PRES: Fred is not a proper part of the sum of all invitees.

            ↙   ↘

       Fred is the sole invitee   ~~Fred was not invited at all~~

A negative cleft, thus, does not presuppose exhaustivity because that particular inference arising from the presupposition is blocked by the assertion in (31-a).

(31) It wasn't Fred she invited.

    a.   ASS: She didn't invite Fred.

    b.   PRES: Fred is not a proper part of the sum of all invitees.

            ↙   ↘

     ~~Fred is the sole invitee~~   Fred was not invited at all

We adopt this analysis for *bá*, make it focus sensitive and add an existence presupposition and a scalar component, see (32). The first line encodes existence, the second exhaustivity, and the third mirativity.

(32) $[\![\text{bá}\ \alpha_F]\ P]\!]^o = \exists x\ [x \in [\![\alpha]\!]^f \wedge [\![P]\!]^o(x) = 1]\ \wedge$
                      $\forall y \in [\![\alpha]\!]^f\ [y \in max([\![P]\!]^o) \rightarrow [\![\alpha]\!]^o \not\sqsubseteq y]\ \wedge$
                      $\forall z \in [\![\alpha]\!]^f\ [z \neq [\![\alpha]\!]^o \rightarrow [\![P]\!]^o(z) \geq_{likely} [\![P]\!]^o([\![\alpha]\!]^o)]\ .$
                      $[\![P]\!]^o([\![\alpha]\!]^o)$

    (where for any $P \in D_{et}, max(P) = \{\ x \in P\ |\ \neg\exists y \in P[x \sqsubset y]\ \}$

This section has developed an analysis for the focus marker *bá* – a particle that operates on focus alternatives, imposing an existence, an exhaustivity, and a mirative restriction on the non-selected alternatives. Evidence for the analysis comes from context tests, co-occurrence with other focus sensitive adverbs, and the projection behaviour.

4   Maximize presupposition

The correction context in (16) only allows for the *bá* strategy. The fact that the *á* strategy is blocked in such contexts straightforwardly follows if we adopt the principle of *maximize presupposition* (Heim 1991, Percus 2006, Sauerland 2008). A definition is given in (33).

(33) *Maximize Presupposition*:
    Do not use ϕ if a member of its Alternative-Family ψ is felicitous and contextually equivalent

to ϕ.

    a. <u>Lexical alternatives</u>: Alternatives are only defined for lexical items. For any lexical item, the alternatives consist of all "presuppositionally stronger" items of the same syntactic category.

    b. <u>Alternative-Family</u>: Let the Alternative-Family of a sentence ϕ be the set of sentences that you get by replacing at least one alternative-associated expression in ϕ with an alternative.

<div align="right">

(Percus 2006)

</div>

The focus particle *bá* belongs to the alternative-family of *á* since it is presuppositionally stronger and of the same syntactic category, i.e. a focus particle that takes a focused constituent as its complement.[6] Hence, it will block *á* in correction contexts where it is felicitous. Other pairs of expression which belong to alternative-families are listed in (34). We would like to add ⟨bá, á⟩ to this list.

(34) ⟨*the, a*⟩, ⟨*both, every*⟩, ⟨*know, believe*⟩, ⟨SING, PLUR⟩, ⟨PAST, PRES⟩, …

The blocking pattern can be replicated for contrast and selection contexts. While the expression of contrast only obligatorily requires a *bá* marker if the selected alternative is truly unexpected, compare (35) to (36), the competing *á* marker is never allowed to occur with the contrasted alternative, see (37) and (38).

(35) Tánkó kí   nɔ̄   mndzīp, Ngàlá cí nɔ̄   (bá) **mbrò?mbvú**
     Tanko HAB drink water   Ngala but drink FOC **palm.wine**
     'Tanko drinks water but Ngala drinks palm wine.'

(36) Tánkó kí   nɔ̄   mndzīp, Ngàlá cí nɔ̄   *(bá) **blēē**
     Tanko HAB drink water   Ngala but drink FOC **blood**
     'Tanko drinks water but Ngala drinks blood.'

(37) *Tánkó kí   nɔ̄   mndzīp, á   **blēē**   cí Ngàlá nɔ̄
     Tanko HAB drink water   FOC **blood** but Ngala drink
     'Tanko drinks water but Ngala drinks blood.'

(38) *Tánkó kí   nɔ̄   mndzīp, á   **mbrò?mbvú** cí Ngàlá nɔ̄
     Tanko HAB drink water   FOC **palm.wine**   but Ngala drink
     'Tanko drinks water but Ngala drinks palm wine.'

Selection contexts, where the alternative set is made explicit, present another environment where we see maximize presupposition at work. A selection context is given in (39), in the form of an alternative question. Note that *bá* cannot occur inside of a coordinate structure, so that the alternative question is in fact ambiguous towards which alternative Shey thinks is less likely. Just as in contrast contexts, *bá* becomes obligatory if the focused constituent expresses an unexpected alternative, which is the case in scenario B but not in scenario A. The *á* strategy, however, is blocked in both scenarios.

---

    [6]Again, see Becker et al. (to appear) for a syntactic analysis.

(39) **Context**: Shey is about to cook dinner. Shey knows that Yaah loves yams and assumes she will prefer it over fufu, but he asks her nevertheless.

Shey: à bí zhē bá **mbrɛ̀ʔ** kɛ̀ **bāā** à
   you FUT2 eat FOC **yams** or **fufu** Q
   'Will you eat fufu or yams?'

A. Shey was right: Yaah prefers yams. Yaah does not know what Shey thinks that Yaah prefers.

Yaah: mɛ̀ bí zhē **mbrɛ̀ʔ**
   I FUT2 eat **yams**
   'I will eat YAMS.'

Yaah′: #á **mbrɛ̀ʔ** (cí) mɛ̀ bí zhē
   FOC **yams** (COMP) I FUT2 eat
   'I will eat YAMS.'

Yaa″: mɛ̀ bí zhē bá **mbrɛ̀ʔ**
   I FUT2 eat FOC **yams**
   'It is yams I will eat.'

B. Shey was wrong: Yaah wants to eat fufu. Yaah knows that Shey knows what Yaah prefers.

Yaah: #mɛ̀ bí zhē **bāā**
   I FUT2 eat **fufu**
   'I will eat FUFU.'

Yaah′: #á **bāā** (cí) mɛ̀ bí zhē
   FOC **fufu** (COMP) I FUT2 eat
   'I will eat FUFU.'

Yaa″: mɛ̀ bí zhē bá **bāā**
   I FUT2 eat FOC **fufu**
   'It is fufu I will eat.'

At this point, it is worth asking whether the unmarked focus strategy can qualify as a competitor for maximize presupposition. Following the definition in (33), it might not since alternatives are only defined for lexical items. However, having the unmarked strategy instantiated as a competitor could potentially explain why it is illicit in (9) on the hand and in (15) and (16) on the other hand, where in the former it loses against the á strategy and in the latter two it loses against the *bá* strategy. The selection and contrast contexts, given in this section, then clearly show that the unmarked strategy is an option so long as mirativity is not established between speaker and hearer. Future research is needed to verify the patterns.

5 Summary

In this paper, we have shown that focus does not need to be morphologically marked in Limbum. The two attested focus particles *á* and *bá* do not encode focus interpretation per se, but rather require focus alternatives to operate on. The focus marker *á* imposes an existence restriction on the alternatives, while the focus marker *bá* additionally (i) exhaustifies over the non-selected alternatives and (ii) restricts the selected alternative to be the least likely. Since we hard-wire mirativity into the semantics of *bá*, unexpectedness as a pragmatic concept is not reflected by the opposition of an unmarked and a marked structure. This is why the *á* strategy can be more marked (particle + fronting) but impose less semantic restrictions, while the *bá* strategy is less marked (only particle) and imposes more semantic restrictions. Limbum, furthermore, provides novel evidence against the assumption that the background, triggered by focus marking, introduces an existence presupposition. At least for Limbum, it can be shown that an additional focus particle is required, i.e. focus marking by itself cannot trigger such a presupposition.

Abbreviations

1,2,3	1st, 2nd, 3rd person	PL	Plural
1-,2-,5-,	Noun classes	PREP	Preposition
COMP	Complementizer	PRV	Preverb
COP	Copula	PST1	Recent past tense
DET	Determiner	PST2	Distant past tense
DEM	Demonstrative	PST3	Remote past tense
EXPL	Expletive	REL	Relative pronoun
FOC	Focus marker	SG	Singular
FUT1	Near future tense		
HAB	Habitual		
INCL	Inclusive		
PERF	Perfective		

References

Alter, K., Mleinek, I., Rohe, T., Steube, A., and Umbach, C. (2001). Kontrastprosodie in Sprachproduktion und -perzeption. In Steube, A. and Umbach, C., editors, *Kontrast - lexikalisch, semantisch, intonatorisch*, volume 77, pages 365–403. Linguistische Arbeitsberichte, Leipzig.

Beaver, D. and Clark, B. (2008). *Sense and Sensitivity: How Focus Determines Meaning*. Explorations in Semantics. Wiley-Blackwell.

Beck, S. (2006). Intervention Effects Follow from Focus Interpretation. *Natural Language Semantics*, 14:1–56.

Becker, L., Driemel, I., and Nformi, J. (to appear). Focus in Limbum. In *Proceedings of ACAL 48*. Language Science Press, Berlin.

Becker, L. and Nformi, J. (2016). Focus and verb doubling in Limbum. *Linguistische Arbeitsberichte*, 93:57–84. Universität Leipzig.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language*, 37:83–96.

Büring, D. (2004). Focus suppositions. *Theoretical Linguistics*, 30:65–76.

Büring, D. and Križ, M. (2013). It's that, and that's it! Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics*, 6:1–29.

Dryer, M. (1996). Focus, pragmatic presupposition, and activated presuppositions. *Journal of Pragmatics*, 26:475–523.

Fiedler, I., Hartmann, K., Reineke, B., Schwarz, A., and Zimmermann, M. (2010). Subject focus in West African languages. In Zimmermann, M. and Féry, C., editors, *Information Structure. Theoretical, Typological, and Experimental Perspectives*, pages 234–257. Oxford University Press, Oxford.

Frey, W. (2010). Ā-Movement and conventional implicatures: About the grammatical encoding of emphasis in German. *Lingua*, 120:1416–1435.

Geurts, B. and van der Sandt, R. (2004). Interpreting focus. *Theoretical Linguistics*, 30:1–44.

Haida, A. (2007). *The Indefiniteness and Focusing of Wh-Words*. PhD thesis, Humboldt University Berlin.

Hartmann, K. (2008). Focus and Emphasis in Tone and Intonational Languages. In Steube, A., editor, *The Discourse-Potential of Underspecified Structures*, pages 389–411. de Gruyter, Berlin/New York.

Hartmann, K. and Zimmermann, M. (2014). In place out of place? focus in hausa. In Schwabe, K. and Winkler, S., editors, *On Information Structure, Meaning and Form: Generalizing Across Languages*, pages 365–403. Benjamins, Amsterdam.

Heim, I. (1991). Artikel und definitheit. In von Stechow, A. and Wunderlich, D., editors, *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, pages 487–535. de Gruyter, Berlin.

Horvath, J. (2010). 'Discourse-Features', Syntactic Displacement and the Status of Contrast. *Lingua*, 120:1346–1369.

Horvath, J. (2013). On focus, exhaustivity and wh-interrogatives: The case of hungarian. In Brandtler, J., Molnar, V., and Platzak, C., editors, *Approaches to Hungarian*, pages 97–132. John Benjamins, Amsterdam.

Katz, J. and Selkirk, E. (2009). Contrastive focus vs. discourse-new: Evidence from prosodic prominence in english. MIT/Umass,Amherst.

Kiss, K. É. (1998). Identification focus versus information focus. *Language*, 74(2):245–273.

Krifka, M. (2008). Basic Notions of Information Structure. *Acta Linguistica Hungarica*, 55(3-4):243–276.

Orenstein, D. and Greenberg, Y. (2013). A flexible exclusive operator: The case of the hebrew *be-sax ha-kol* ($\approx$ *all in all*). In Chemla, E., Homer, V., and Winterstein, G., editors, *Proceedings of Sinn und Bedeutung 17*, pages 381–397, Paris.

Percus, O. (1997). Prying open the cleft. In Kusumoto, K., editor, *Proceedings of NELS 27*, pages 337–351. McGill University.

Percus, O. (2006). Antipresuppositions. In Ueyama, A., editor, *Theoretical and Empirical Studies of Reference and Anaphora: toward the Establishment of Generative Grammar as an Empirical Science*, page 5273.

Rooth, M. (1985). *Association with Focus*. PhD thesis, University of Massachussetts.

Rooth, M. (1992). A Theory of Focus Interpretation. *Natural Language Semantics*, 1:75–116.

Rooth, M. (1999). Association with focus or association with presupposition? In Bosch, P. and van der Sandt, R., editors, *Focus*, Studies in Natural Language Processing, pages 232–244. Cambridge University Press, Cambridge.

Sauerland, U. (2008). Implicated presuppositions. In Steube, A., editor, *The discourse potential of underspecified structures*, pages 581–600. de Gruyter, Berlin.

Skopeteas, S. and Fanselow, G. (2009). Focus types and argument asymmetries: a cross-linguistic study in language production. In Breul, C. and Göbbel, E., editors, *Contrastive Information Structure*, pages 169–198. Benjamins, Amsterdam/Philadelphia.

Skopeteas, S. and Fanselow, G. (2011). Focus and the exclusion of alternatives: On the interaction of syntactic structure with pragmatic inference. *Lingua*, 121:1693–1706.

Szabolcsi, A. (1981). Compositionality in focus. *Folia Linguistica Societatis Linguisticae Europaeae*, 15:141–162.

Vallduví, E. and Vilkuna, M. (1998). On Rheme and Kontrast. In Culicover, P. and McNally, L., editors, *The Limits of Syntax*, pages 79–106. Academic Press, New York.

van der Wal, J. (2011). Focus excluding alternatives: Conjoint/disjoint marking in Makhuwa. *Lingua*, 121:1734–1750.

van der Wal, J. (2014). Subordinate clauses and exclusive focus in Makhuwa. In *Information Structure and Reference Tracking in Complex Sentences*, Typological Studies in Language 105, pages 45–70. John Benjamins, Amsterdam.

Velleman, D., Beaver, D., Destruel, E., Bumford, D., Onea, E., and Coppock, L. (2012). It-clefts are IT (Inquiry Terminating) constructions. In Chereches, A., editor, *Proceedings of SALT 22*, pages 441–460. University of Chicago.

Zimmermann, M. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica*, 55:347–360.

Zimmermann, M. (2011). The grammatical expression of focus in West Chadic: Variation and uniformity in and across languages. *Linguistics*, 49:1163–1213.

Zimmermann, M. and Onea, E. (2011). Focus marking and focus interpretation. *Lingua*, 121:1651–1670.

Intensifying ideophones in three Luhya languages[1]

Margit Bowler — University of California, Los Angeles
John Gluckman — University of California, Los Angeles

Abstract. Ideophones are typically described as "marked words that depict sensory imagery" (Dingemanse 2011, 25). This paper addresses ideophone data from three Luhya languages: Llogoori, Lunyore, and Lutiriki (Bantu, Kenya). Our primary claim is descriptive: we show that there is a closed class of (previously undescribed) Luhya ideophones. We illustrate how the Luhya data is consistent with what is known about ideophones cross-linguistically, and give a preliminary semantic analysis of the Luhya ideophones as degree intensifiers.

1   Introduction

This paper addresses ideophones in three Bantu languages in the Luhya subfamily: Llogoori, Lunyore, and Lutiriki.[2] In this paper, we make both descriptive and theoretical contributions to the ideophone literature. Descriptively, we add novel data to the existing typology of ideophones cross-linguistically. We illustrate that, despite looking superficially distinct from other documented ideophone systems, the Luhya ideophones have the core properties of ideophones. Theoretically, we give a preliminary semantic analysis of the Luhya ideophones as degree intensifiers.

Ideophones have been described across the world, including in the languages of Asia (Japanese; Hamano 1994), Australia (Yir-Yoront; Alpher 1994), the Americas (Tseltal; Henderson 2016), and Europe (Basque; Antuñano 2016). Some authors, including Voeltz and Kilian-Hatz (2001), argue that ideophones occur in every language. Despite their frequency, the debate for how to classify a given lexical item as an ideophone is far from settled. To this end, we begin by reviewing the existing ideophone literature to give a general definition for what makes a lexical item an ideophone.

1.1   How to classify a lexical item as an ideophone

Ideophones are lexical items that often describe sensory imagery and tend to be morphosyntactically "marked" in some way (Dingemanse 2011, Voeltz and Kilian-Hatz 2001, Childs 1994, Doke

---

[2]The Luhya subfamily (Guthrie: JE.41, JE.30, JE.18) consists of 25 (or so) closely related languages spoken in western Kenya, northwestern Tanzania, and eastern Uganda. There are approximately 5 million speakers of Luhya languages, with a relatively high degree of mutual comprehension between speakers of different languages (Simons and Fennig 2017, Marlo 2017). Llogoori is also referred to as Maragoli, Luragooli, and Logoori, among other names; Lutiriki is also referred to as Tiriki.

1935, among many others).[3] Dingemanse (2012, 654) remarks that ideophones are "easy to identify, but difficult to define;" typological work has shown that ideophones have a wide range of phonological and morphosyntactic properties. We therefore begin by giving examples of typologically diverse ideophone data from Kisi (Niger-Congo), Wolaitta (Omotic), Tseltal (Mayan), and Tsonga (Bantu).

(1) **Kisi (Niger-Congo)**

ò kwé **déèè**...
PRO go IDEO

'She went *déèè* (slowly).'[4]   (Childs 1988, 178-179)

(2) **Wolaitta (Omotic)**

Galláso-y **k'ap'k'áp'a**.
Gallasso-NOM IDEO

'Gallasso is *k'ap'k'áp'a* (greedy).'   (Amha 2001, 57)

(3) **Tseltal (Mayan)**

pura ch'il-bil-∅, **tsok'** x-chi-∅ ta mantekat.
just fried-PERF-B3 IDEO NT-say-B3 P lard

'Just fried, it goes *tsok'* in the lard.'   (Henderson 2016, from Pérez González 2012, 162)

(4) **Xitsonga (Bantu)**

Magezi u ri ti-**nka**, hi xihloka.
Magezi SC1 COP REFL-IDEO by axe

'Magezi chops himself *nka* with an axe.'

                               (Msimang and Poulos 2001, 240, from Marivate 1982)

The data in (1)-(4) demonstrates some of the grammatical properties that are often (but not always) described for ideophone systems cross-linguistically. Common phonological properties of ideophones include: (i) the ability to lengthen vowels for expressive effect, as in (1); (ii) the ability to partially or totally reduplicate the ideophone, as in (2); (iii) the presence of sound symbolism or onomatopoeia, perhaps as in (3); (iv) the presence of phonemes or tones not otherwise found in the language; and (v) unusual phonation such as creaky voice, breathy voice, or falsetto. These unusual phonological and phonetic properties have led linguists to propose that ideophones often contribute meaning that is "depictive rather than descriptive" (Essegbey 2013).

Common morphosyntactic properties of ideophones include: (i) occurring clause-peripherally, as in (1) and (2); (ii) co-occurring with a quotative marker or verb of saying or doing, as in (3); (iii)

---

[3]Here and elsewhere in this paper, we do not use the term "expressive" in the Pottsian sense. Instead, we follow the convention in the ideophone literature to use it as a conceptual description of the ideophones' often onomatopoeic or "depictive" properties.

[4]Abbreviations used in this paper include: 1-20 'noun class,' 1 'first person,' 2 'second person,' 3 'third person,' AC 'anticausative,' ASP 'aspect,' AUX 'auxiliary,' CAUS 'causative,' CL 'noun class,' COP 'copula,' DEF 'definite,' EXPR 'expressive,' FUT 'future,' FV 'final vowel,' IDEO 'ideophone,' NEG 'negation,' NOM 'nominative,' PERF 'perfective,' PL 'plural,' POSS 'possessive,' PRO 'pronoun,' PROG 'progressive,' PRT 'particle,' QM 'quotative marker,' REC 'reciprocal,' REFL 'reflexive,' SC 'subject class,' SG 'singular,' TNS 'tense.'

the inability to combine with other morphemes, as contradicted by (4); (iv) patterning morphosyntactically distinctly from other lexical categories in the language; and (v) the ability to stand alone as a complete utterance.

As shown by (1)-(4), there is a great deal of variation in ideophone systems cross-linguistically. We now review Dingemanse and Akita (2016)'s proposed ideophone typology, as this range of variation is relevant to our claim that the Luhya lexical items we discuss are in fact ideophones.

1.1.1 Variation in ideophone systems: Dingemanse (2017) and Dingemanse and Akita (2016)

Dingemanse (2017) and Dingemanse and Akita (2016) argue that ideophones occur along inversely correlated scales of "expressiveness" and "grammatical integration."[5] Their criteria for expressiveness and grammatical integration are based on existing typological observations about ideophone systems, as laid out in §1.1.

Dingemanse and Akita (2016) propose that an "expressive" ideophone shows some or all of the following properties: (i) intonational foregrounding through marked prosody, lengthened vowels, or so on; (ii) unusual phonation; (iii) the presence of tones or phonemes not found elsewhere in the language; and (iv) accompaniment by iconic gesture. A "grammatically integrated" ideophone shows some or all of the following properties: (i) inability to stand alone as a complete utterance; (ii) ability to occur clause-internally; (iii) ability to embed in morphosyntactic structure; and (iv) lack of syntactic optionality. These properties are summarized in Figure 1 below.

Bound constructions   Free constructions
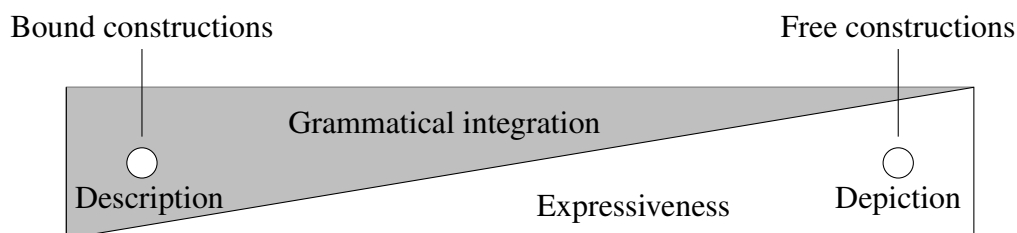
Figure 1: Inverse correlation between grammatical and expressive properties of ideophones (Dingemanse 2017, 133).

Dingemanse (2017) proposes that ideophone systems can vary with respect to their expressiveness versus grammatical integration. For instance, Dingemanse observes that ideophones in Semai (Mon-Khmer) tend to be highly expressive, whereas ideophones in Somali (Cushitic) tend to be more grammatically integrated. However, internal variation within a single language's ideophone system is also possible: one language may have both expressive and grammatically integrated ideophones. Indeed, Dingemanse (2017) argues that in Siwu (Niger-Congo), a single ideophone can be more or less expressive in different contexts.

In §2, we show ideophone data from Llogoori, Lunyore, and Lutiriki. We argue that (in Dingemanse and Akita 2016's terms) the Luhya ideophones are relatively highly grammatically integrated, with a corresponding relatively low degree of expressiveness.

---

[5]Dwyer and Moshi (2003) make a similar observation about ideophone classes. They propose to distinguish between "primary" ideophones (corresponding roughly to Dingemanse and Akita 2016's expressive ideophones) and "grammaticalized" ideophones (corresponding roughly to Dingemanse and Akita 2016's grammatically integrated ideophones).

2   Basic Luhya ideophone data

We give examples of Llogoori, Lunyore, and Lutiriki ideophones in (5)-(7). (Hereafter, unless otherwise noted, we give examples in Llogoori, our primary language of study.) The Luhya ideophones select for a semantic class of lexical items that they can co-occur with. Luhya ideophones typically occur clause-finally and provide an "intensified" reading of the lexical item that they select. As suggested by (5)-(7), the Luhya ideophones tend to be cognate across the languages.

(5)   **Llogoori**

    a. amaaze ni   ma-hiu **pa**.
       6.water COP 6-hot   IDEO
       'The water is very hot.'

    b. riawa   ni   ri-akanyu **khai**.
       5.flower COP 5-red   IDEO
       'The flower is very red.'[6]

(6)   **Lunyore**

    a. maatsi ne   ma-hiu **pa**.
       6.water COP 6-hot   IDEO
       'The water is very hot.'

    b. esausi ne   i-nzakanyu **kha**.
       9.sauce COP 9-red   IDEO
       'The sauce is very red.'

(7)   **Lutiriki**

    a. matse ni   ma-hiu **pa**.
       6.water COP 6-hot   IDEO
       'The water is very hot.'

    b. intso   ni   y-amuchi **kha**.
       9.house COP 9-red   IDEO
       'The house is very red.'

We list some Llogoori ideophones and their associated semantic classes in Table 1. The items within a given semantic class all have similar meanings; for instance, the ideophone *ti* combines with lexical items describing darkness or dirtiness, whereas *zi* combines with lexical items describing stillness or coldness. (The data in Table 1 is not an exhaustive list of all the lexical items each ideophone can co-occur with, nor is it an exhaustive list of all of the Llogoori ideophones.)

Ideophone	Lexical item(s)[7]	Meaning
mno[8]	*kuyaanza* (verb)	'to be happy,' 'to like'
	mahooru (noun)	'longing'
	-ndugi, -noru (adjective)	'sweet'
pa	*-hiu* (adjective)	'hot'
	kuhia (verb)	'to be hot'
	-roro (adjective)	'spicy,' 'bitter'
ti	*-mwamu* (adjective)	'black'
	-chafu (adjective)	'dirty'
zi	*-zilu* (adjective)	'cold,' 'still'
	-chinganu (adjective)	'quiet'

Table 1: Lexical items selected by Llogoori ideophones.

---

[6]The voiceless velar fricative *kh* in (5b) is an uncommon phoneme in Llogoori, although it is frequent in many of the other closely related Luhya languages.

The Luhya ideophones cannot occur with lexical items outside of the semantic class that they select. For instance, the Llogoori ideophones *pa* and *khai* in (5) cannot be substituted for the other, as in (8). The ideophone *pa* is restricted to lexical items describing hotness or spiciness, whereas *khai* is restricted to lexical items describing redness. Furthermore, ideophones can pick out only a subset of meanings within their given semantic class. The ideophone *du* can occur with lexical items describing fullness in the sense of a cup or a room, as in (9a); however, it cannot occur with an expression describing a person's sensation of being full, as in (9b).

(8) a. * amaaze ni   ma-hiu khai.
     6.water COP 6-hot   IDEO
     Intended: 'The water is very hot.'

    b. * riawa   ni   ri-akanyu pa.
     5.flower COP 5-red   IDEO
     Intended: 'The flower is very red.'

(9) a. kikoombe ki-ikwizor-a du.
     7.cup   7-full-FV   IDEO
     'The cup is very full.'

    b. * Sira y-a-ku-i-goot-a   du.
     Sira 1-TNS-ASP-REFL-sate-FV IDEO
     Intended: 'Sira is very full.'

This property of semantic class selection distinguishes the Luhya ideophones from the Luhya degree intensifier *saana* 'really.' We show Luhya degree intensifier data, and discuss how *saana* 'really' differs from the ideophones, in §3.

2.1 Grammatical properties of the Luhya ideophones

Luhya ideophones can occur with adjectival predicates, as in (5), and with verbal predicates, as in (10)-(11).[9]

---

[7]We give verbs in their infinitival form, including the class 15 infinitival prefix *ku-*. We give adjectives in their root form; Luhya adjectives obligatorily host a prefix indicating the noun class of the noun that they combine with.

[8]Mike Marlo (p.c.) notes that in Swahili, *muno* is a canonical degree intensifier (like English *really*) that is not restricted to any lexical class. While Llogoori *mno* is likely a borrowing from Swahili, its distribution differs from Swahili in that it is in fact subject to lexical restrictions. This could be a point of variation across Luhya; in Lutiriki, *mno* appears to pattern more like Swahili.

[9]Luhya ideophones almost always combine with stative predicates. However, a very small number of ideophones given to us by our Lunyore consultant can combine with eventive predicates:

(1) **Lunyore**

    a. esaal'a si-mekukh-il-e   piap.
     9.stick 9-break-TNS-FV IDEO

     'The stick broke *piap*.'

    b. ya-khu-pak-il-e   pap.
     1-ASP-hit-TNS-FV IDEO

     'He just hit me *pap*.'

Unlike the other Luhya ideophones that we've discovered, our speaker reported that *piap/pap* is the sound that breaking and hitting make; that is, they are iconic. These ideophones, like the others, are limited to combining with a particular semantic class: *pap* can only describe a hitting event, whereas *piap* can only describe a breaking event.

We ultimately choose to exclude these ideophones from our analysis. We suspect that they are borrowings from a Luo language; our Lunyore consultant also speaks fluent Luo, is married to a Luo speaker, and regularly uses Luo in her daily life. Furthermore, these ideophones resemble typical Nilotic ideophone data (Mark Dingemanse, p.c.).

(10) **Llogoori**
Sira   yi-zuriz-i kikoombe du.
1.Sira 1-fill-FV 7.cup   IDEO
'Sira filled the cup to the brim.'

(11) **Lunyore**
rishirti ri-n-nyik-il-e   ka.
5.shirt 5-1SG-be.tight-APPL-FV IDEO
'The shirt is very tight (on me).'

A small number of Luhya ideophones can combine with nouns, as in (12)-(13).[10]

(12)   a. inzankanyu khai
          9.redness   IDEO
          'intense redness'

       b. uvwizulu du
          11.fullness IDEO
          'extreme fullness'

       c. mahooru mno
          6.longing IDEO
          'intense longing'

(13)   m-v-ey-e   na mahooru mno.
       1SG-COP-ASP-FV NA 6.longing IDEO
       'I really miss you.'
       (Lit. 'I am with intense longing.')

The Luhya ideophones cannot stand alone as predicates; that is, they cannot occur without any associated lexical item, as in (14).[11]

(14)   amaaze ni   *(ma-hiu) pa.
       6.water COP 6-hot   IDEO
       Intended: 'The water is very hot.'

The ideophones also cannot stand alone as complete utterances, as in (15)-(16). (We return to this issue in §3.2.3.)

(15) **Hot bathwater context:** You run a bath, then touch the bathwater and discover that it's extremely hot. You exclaim:

   a. * pa!
          IDEO

   b. ɬːↆ̣ha̤!
       EXPR
       'Ouch!'[12]

---

[10]We previously postulated that the ideophones can combine with "prepositional predicates," consisting of the preposition *na* 'with' followed by the noun and ideophone, as in (13). However, given the new data in (12), we now assume they combine directly with the noun, which then in turn can combine with *na*.

[11]A possible exception to this is the ideophone *du*, which one Llogoori consultant accepts as a predicate. Curiously, this ideophone is only accepted in combination with the copula *kova*, and not the copula *ni*. We currently have no explanation for these facts, and do not account for them in our analysis in §3.

(1)   a. kikoombe ki-v-ey-e   du.
         7.cup   7-COP-ASP-FV IDEO
         'The cup is full.'

      b. * kikoombe ni   du.
           7.cup   COP IDEO
           'The cup is full.'

[12]Here we use the Extended IPA symbol ↆ to represent ingressive airflow during the production of the lateral fricative. So far, we have collected approximately 15 Llogoori (Pottsian) expressives that are akin to English expressives like *ouch* and *oops*. These morphemes pattern syntactically very differently from the Luhya ideophones; they can stand alone as complete utterances, and they necessarily precede the proposition they co-occur with.

(16) **Sweet tea context:** Imali makes you some tea and asks how sweet it is with the question in (16a). You respond as in (16b).

    a. icha i-v-ey-e   na uvunoru   vuri?
        9.tea 9-COP-ASP-FV with 11.sweetness how.much
        'How sweet is the tea?'

    b. i. * mno!   ii. saana!   iii. ni i-noru mno!
             IDEO   really   COP 9-sweet IDEO
                                      'Very!'   'It is very sweet!'

The Luhya ideophones cannot be moved away from their associated lexical item, as the cleft construction in (17) illustrates, and do not combine with complementizers, quotative markers, or light verbs of saying or doing, as in (18).

(17)   * du   ni   sia Sira y-izuriz-i kikoombe.
          IDEO COP how 1.Sira 1-fill-FV 7.cup
          Intended: 'To the brim is how Sira filled the cup.'

(18)   * maaze ni   ma-hiu {kuresia / ndee / ga-vor-a} pa.
          6.water COP 6-hot {like   / COMP / 6-say-FV} IDEO
          Intended: 'The water is hot like *pa*.' / 'The water goes *pa*.'

The Luhya ideophones behave in many ways like adverbial elements; they typically occur at the right edge of the clause, and are always syntactically optional, as in (19). In expressions with verbal predicates, ideophones occur immediately after the direct object, inside of other verbal modifiers such as manner adverbs, as in (20). In the presence of an applied object, such ideophones occur after both objects, as in (21).

(19)   (*pa) kibiribiri ni   (*pa) ki-roro (pa).
        IDEO 7.pepper COP IDEO 7-spicy IDEO
        'The pepper is (very) spicy.'

(20)   Imali yi-zuriz-i (*du) kikoombe (du)   {geraha / na   maaze} (*du).
        Imali 1-fill-FV (IDEO) 7.cup   (IDEO) {slowly / with 6.water} (IDEO)
        'Imali filled the cup to the brim slowly/with water.'

(21)   Imali yi-zuriz-il-i   (*du) Sira (*du) kikoombe (du).
        Imali 1-fill-APPL-FV (IDEO) Sira (IDEO) 7.cup   (IDEO)
        'Imali filled the cup for Sira to the brim.'

If the adjective that the ideophone selects for is in an attributive position, the ideophone must occur immediately after it; it cannot occur at the end of the clause, as in (22b). (We are agnostic as to whether Luhya attributive adjectives involve relative clauses.)

(22)   a.   maaze ma-hiu pa   ga-v-ey-e   mu kikoombe.
            6.water 6-hot   IDEO 6-COP-ASP-FV in 7.cup
            'The very hot water is in the cup.'

b. * maaze ma-hiu ga-v-ey-e   mu kikoombe pa.
  6.water 6-hot 6-COP-ASP-FV in 7.cup   IDEO
  Intended: 'The very hot water is in the cup.'

If multiple adjectives modify a single noun, the order of the adjectives is free. However, the ideophone must immediately follow the adjective that it selects for.

(23) a. riaua   ri-nini ri-akanyu khai
     5.flower 5-big 5-red   IDEO
     'the big very red flower'

   b. riaua   ri-akanyu khai ri-nini
     5.flower 5-red   IDEO 5-big
     'the big very red flower'

   c. * riaua   ri-akanyu ri-nini khai
     5.flower 5-red   5-big IDEO
     Intended: 'the big very red flower'

All of the data so far is generally consistent with treating the Luhya ideophones as adverbs. However, we argue in §2.3 that the Luhya ideophones should be treated as a unique class, distinct from adverbs.

2.2 Expressiveness of the Luhya ideophones

All Luhya ideophones can be reduplicated to express a more intense meaning.[13] A small set of ideophones can also undergo triplication, which results in a further intensified meaning and occurs with a unique prosodic contour that is used for all triplicated ideophones. (We return to the triplicated ideophones in §3.2.3.)

(24) amaaze ni   ma-hiu pa   pa.   (25) amaaze ni   ma-hiu papapa.
    6.water COP 6-hot   IDEO IDEO   6.water COP 6-hot   PAPAPA
    'The water is extremely hot!'   'The water is BOILING hot!'

The Luhya ideophones are not inherently associated with either negative or positive evaluations. That is, the utterance in (26) is felicitous in a context in which the water being very hot is a good thing (26a), a bad thing (26b), or neither.

(26) a. **Positive evaluation context:** You fill a bath for your wife, who is cold and wants to warm up. You tell her that the water is ready (i.e., it's very hot).

   b. **Negative evaluation context:** You try to enter a hot tub and discover that the water is too hot to be comfortable. You warn another nearby bather about its temperature.

---

[13]The sole trisyllabic ideophone in Llogoori, *zululia*, expresses this additionally intensified meaning by lengthening the second vowel: /zulu:lia/. (This ideophone combines with predicates describing vertical height, e.g. *-tambe* 'tall'.)

amaaze ni   ma-hiu pa.
6.water COP 6-hot   IDEO
'The water is very hot.'

The Luhya ideophones only occasionally display marked phonation or intonation (typically a raised pitch), and are only occasionally accompanied by iconic gestures (Mike Marlo, p.c.). They are also able to be used naturally in written language. Speakers do not report that the Luhya ideophones are interpreted iconically; one possible exception to this is *pa*, which may be interpreted as the sound of water boiling. Finally, the Luhya ideophones are not "productive;" speakers cannot spontaneously coin new ones, unlike reports of spontaneous ideophone generation in languages like Semai (Mon-Khmer) (Diffloth 1972). The Luhya ideophones form a closed class; we have identified fewer than 15 ideophones in each of the languages that we investigated.

2.3   Why do we call these morphemes ideophones?

In the terminology of Dingemanse (2017) and Dingemanse and Akita (2016), the Luhya ideophones display a relatively low degree of expressiveness and a relatively high degree of grammatical integration, as shown in Figure 2. (Compare with Figure 1.) The Luhya languages pattern similarly to languages like Somali (Cushitic) in having consistently "non-expressive" ideophones (Dhoorre and Tosco 1998).
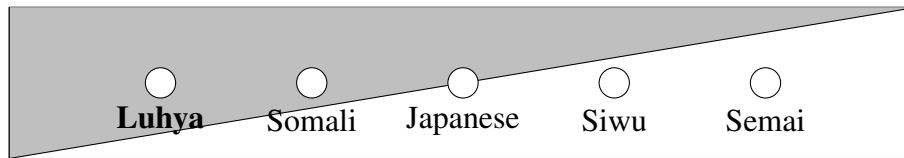
Figure 2: Approximation of different ideophone systems on a scale of grammatical integration/expressiveness; adapted from Dingemanse (2017, 136).

Given the lack of expressiveness and their syntactic similarity to other adverbial items, the reader may question why we choose to call these lexical items ideophones. However, like other described ideophone systems, the Luhya ideophones: (i) are constrained in their distribution by the semantic class of the lexical item they select for; (ii) undergo reduplication to express additional intensification; (iii) (almost always) have a fixed number of syllables; (iv) have a consistent syllable shape (i.e., they are almost always open syllables); (v) can contain phonemes that are otherwise infrequent in the languages; and (vi) denote "exaggerated" shades of meaning (i.e., *extremely hot*, not *lukewarm*).

Furthermore, a number of other languages have similar ideophone systems. We give examples in (27)-(31) of intensifying ideophones in Hausa (Chadic), Siwu (Niger-Congo), Wolof (Niger-Congo), Xitsonga (Bantu), and Zulu (Bantu). In each of these examples, the relevant ideophone patterns like the Luhya ideophones in (i) selecting for a particular lexical item or semantic class, and (ii) contributing an intensified reading of the lexical item that it selects for. Note that several of these languages are typically thought of as having canonical examples of highly depictive ideophone systems.

(27) **Hausa (Chadic)**

a. fari **fat**
   white IDEO
   'snow white'[14]

b. tsofo **kutuf**
   old IDEO
   'very old'   (Newman 1968, 109)

(28) **Siwu (Niger-Congo)**

i-tì   si i-fudza-ɔ   ↑**fututututututu**↑.
C.I-head if S.I-be.white-2SG.O IDEO.pure.white.EM4

'That your head may become white ↑*fututututututu*↑ [pure white].'[15]

(Dingemanse 2017, 123)

(29) **Wolof (Niger-Congo)**

daf-a   ñuul **kukk**.
do-COP black IDEO

'It's pitch black.'   (Harold Torrence, p.c.)

(30) **Xitsonga (Bantu)**

khuwani ri   tele   **ntlwi**!
clay.pot COP be.full IDEO

'The clay pot is very full.'   (Kubayi 2009, 43)

(31) **Zulu (Bantu)**

w-a-thula   **du**.
1SG-PST-be.silent IDEO

'(S)he was absolutely silent!'   (Claire Halpert, p.c.)

Furthermore, we observe that there are ideophones in many non-Luhya Bantu languages that are cognate with the Luhya ideophones (Samarin 1971). Indeed, previous classifications of similar items in other Bantu languages directly refer to these lexical items as ideophones (Schadeberg 2003). Thus, given the previous classification, the ideophonic properties of the relevant Luhya lexical items, and the existence of other intensifying ideophones cross-linguistically, we feel justified in our proposal to treat these items as ideophones.

3   Towards an analysis

Since Luhya ideophones intensify the predicate that they combine with, we propose to treat them as degree modifiers akin to English *very* or *really*. In the following section, we lay out our proposal to treat the Luhya ideophones as cross-categorial degree modifiers. We begin by providing a brief background on degree semantics.

---

[14]We note that a small set of English adjectives combine with similarly lexically restricted intensifiers; these include *jet black* and *bitter cold*, among others. However, English expressions like *jet* and *bitter* occur elsewhere in the language as fully fledged lexical items. We therefore do not believe that these lexical items should be considered ideophones.

[15]Dingemanse (2017) uses arrows ↑ to indicate general prosodic foregrounding.

3.1 Degrees

Degree theories of gradable adjectives argue that gradable predicates combine with both a degree argument ($d \in D_d$) and an individual argument ($x \in D_e$), and assert that the adjective holds of the individual x to degree d (Bartsch and Vennemann 1972, Cresswell 1976, Heim 2001, among many others). We give a basic denotation for the English gradable adjective *hot* in (32).

(32) ⟦hot⟧ = λdλx. hot(x,d)   ("x is hot to degree d")

English degree intensifiers like *really*, *very*, *extremely*, and so on contribute the meaning that the degree of the adjective with respect to the individual is above some contextual standard.[16] We give a basic denotation for the English degree intensifier *really* in (33a), and provide a denotation for the intensified adjective *really hot* in (33b).

(33) a. ⟦really⟧ = λG$_{<d<et>>}$λx. ∃d: G(x,d) & d > standard
      b. ⟦really hot⟧ = λx. ∃d: hot(x,d) & d > standard

          ("there exists a degree *d* such that *x* is hot to degree *d* and *d* exceeds the
           contextual standard of hotness")

3.2 Luhya ideophones as cross-categorial degree intensifiers

We propose that the Luhya ideophones, like English *really*, are fundamentally degree intensifiers. The ideophones provide an extremely intensified reading of the predicate that they combine with. In combination with gradable adjectives, the ideophones assert that the degree to which the gradable adjective holds greatly exceeds the contextual standard (represented in (34) with "!!>").[17] We give the truth conditions for a Llogoori expression containing the ideophone *pa* in (35).

(34) a. ⟦IDEO⟧ = λG$_{<d<et>>}$λx. ∃d: G(x,d) & d !!> standard
      b. ⟦hot IDEO⟧ = λx. ∃d: hot(x,d) & d !!> standard

(35) maaze ni   ma-hiu pa.
      6.water COP 6-hot   IDEO
      'The water is very hot.'   = 1 iff ∃d: hot(water,d) & d !!> standard

The Luhya ideophones differ from the canonical Luhya degree intensifier *saana* 'really' in three main ways. First, speakers report that *saana* contributes a less intensified reading of the predicate that it combines with. Second, *saana* is not restricted to combining with any particular semantic class; it freely combines with all gradable predicates. Third, *saana* can stand alone as a felicitous answer to degree questions (16b-ii), whereas ideophones cannot (16b-i).

This analysis of the ideophones as degree intensifiers can easily account for their ability to co-occur with adjectival predicates, as shown in (5)-(7). All of the adjectives that the ideophones

---

[16]In utterances without any degree intensifier or measure phrase, we assume that the predicate combines with some phonologically null morpheme that contributes the meaning that the individual that the predicate combines with "stands out" with respect to the property denoted by the predicate (Kennedy 1999, Rett 2008). We remain agnostic with respect to the precise denotation for this morpheme, since it is not crucial to our proposal.

[17]This notation is inspired by Kennedy and McNally (2005, 373)'s proposal for English *much*.

co-occur with are uncontroversially associated with degree scales (e.g. *-hiu* 'hot,' *-zilu* 'cold,' *-noru* 'sweet,' and so on).[18] In the following subsections, we address how we can extend this proposal to account for the ability of the ideophones to combine with nominals and verbal predicates.

3.2.1 Ideophones in combination with nominals

Only three of the Luhya ideophones can combine with nominals in addition to adjectives. We show examples of these three ideophones in (36), repeated from (12).

(36) a. inzankanyu khai   b. uvwizulu du   c. mahooru mno
      9.redness   IDEO   11.fullness IDEO   6.longing IDEO
      'intense redness'   'extreme fullness'   'intense longing'

To account for this data, we assume that a subset of Luhya nouns (*mahooru* '6.longing,' *vuyaanzi* '11.happiness,' and *uvwakanyu* '11.redness,' among others) include degrees in their denotations. Proposing to introduce degrees into the nominal domain is not novel; several authors have previously argued that some nouns include degrees (Morzycki 2009 for English, Bochnak 2013 for Luganda, among others).

We roughly assume Bochnak (2013)'s analysis of verbal nominalizations in Luganda (Bantu). Bochnak proposes that Luganda nominalized gradable predicates are relational: that is, they denote relations between individuals and degrees. (This follows prior proposals for relational nouns by Nicolas 2004 and Moltmann 2009.) We assume this analysis for the relevant Luhya nominals, which we also term "relational." However, we note that the Luhya relational nouns differ from Bochnak (2013)'s Luganda nominals in that they do not have a verbal core. The Luhya relational nouns are of type $<e<d,t>>$; we use the variable R to refer to items of this type. We give a denotation for the Llogoori relational noun *mahooru* '6.longing' in (37).

(37) ⟦mahooru⟧ = $\lambda x \lambda d.$ longing(x,d)   "the individual x instantiates longing to degree d"

To account for the ability of ideophones to combine with the relational nouns, we assume that the ideophones in (36) have the denotation in (38a), termed IDEO$_N$. This is identical to the denotation for ideophones that combine with adjectives, as in (34a), with the exception of the semantic type of the first argument that the ideophone combines with.

(38) a. ⟦IDEO$_N$⟧ = $\lambda R_{<e<dt>>} \lambda x. \exists d: R(x,d)$ & d !!>
     b. ⟦mahooru IDEO$_N$⟧ = $\lambda x. \exists d:$ longing(x,d) & d !!> standard

To account for data like (39), we assume that *na* is of type $<<e,t><e,t>>$.

---

[18]An apparent possible counterexample to this is the ability of the ideophone *zi* to combine with the verb *kukuzila* 'to be cold'/colloquially, 'to be dead.' In these cases, we assume the gradable meaning of 'to be cold' as the basic meaning of *kukuzila*. Our Llogoori consultant reports that such uses of *zi* are a colloquialism meaning 'to be dead.'

(1) imbwa y-a-kuzil-a   zi.
     9.dog   9-TNS-cold-FV IDEO

     'The dog is dead as a doornail.'

(39) m-v-ey-e   na mahooru mno.
    1SG-COP-ASP-FV NA 6.longing IDEO
    'I really miss you.'   (Lit. 'I am with intense longing.')
    = 1 iff ∃d: longing(I,d) & d !!> standard

Postulating a second denotation for the Luhya ideophones introduces a bit of messiness into our analysis. However, the ambiguity proposed here is consistent with the overall distribution and use of ideophones cross-linguistically. As we discussed in §1.1.1, following Dingemanse (2017) and Dingemanse and Akita (2016), one language may have many different kinds of ideophones. Given what we know about the diversity of ideophone systems both within and across languages, we have no reason to assume that the Luhya ideophones should form a homogenous class with respect to their semantics.

3.2.2 Ideophones in combination with verbal predicates

We observed in §2.1 that the Luhya ideophones can also combine with verbal predicates, as in (40)-(41).

(40) **Lutiriki**

Sira   yi-tsurits-a shikoombe tu.
1.Sira 1-fill-FV   7.cup   IDEO

'Sira filled the cup to the brim.'

(41) **Llogoori**

marova ga-uum-i gada.
6.earth 6-dry-FV IDEO

'The earth dried a lot [until it was hard].'

All of the verbs that can co-occur with ideophones have a gradable adjectival core, including *kumwama* 'to blacken' (from -*mwamu* 'black'), *kwuuma* 'to dry' (from -*uumu* 'dry'), and so on. As shown in (34), it is simple to treat the ideophones as degree modifiers of gradable adjectives. Although we do not give a full semantics for the Luhya ideophones in combination with verbs in this paper, we believe that the gradable adjectival core of these verbs can provide a starting point as to their semantics.

Kennedy and Levin (2008) give a semantics for English degree achievement verbs (e.g. *to cool*, *to widen*) that uses degrees. They link the use of degrees in the semantics to the gradability of the verbs' adjectival cores. They propose that degree achievement verbs include a derived measure of change function that measures the degree to which an object changes along a scalar dimension as the result of participating in an event.[19] We set aside the precise formal implementation of this theory for now; however, we note that a proposal along these lines that either includes or introduces degrees in the denotations of verbs derived from gradable adjectives could account for the data in (40)-(41). If we follow Kennedy and Levin (2008)'s proposal, the paraphrased meaning of (40),

---

[19] This measure of change function $m_\triangle$ is defined formally as follows (Kennedy and Levin 2008, 18):

(1)   For any measure function m, $m_\triangle = \lambda x \lambda e.\ m\uparrow_{m(x)(init(e))}(x)(fin(e))$,
      where *init(e)* and *fin(e)* refer to the initial and final temporal intervals of an event, and $m\uparrow_d$ is a difference function that takes an individual and returns the difference between the individual's projection on a degree scale and the (arbitrary) comparative standard.

including the ideophone, would be something like "Sira filled the cup to a degree that greatly exceeds the contextual standard of what counts as 'full.'"

3.2.3 Triplicated Luhya ideophones

Some of the Luhya ideophones can undergo triplication. When triplicated, the ideophones pattern very differently from non-triplicated ideophones. The triplicated ideophones can be clefted (contrary to (17)), and can stand alone as complete utterances (contrary to (15)-(16)).[20]

(42) dududu   ni   sia Sira   y-izuriz-i kikoombe.
      DUDUDU COP how 1.Sira 1-fill-FV 7.cup
      '*Dududu* (to the brim) is how Sira filled the cup.'

(43) Sira y-izuriz-i kikoombe ndi   nang'ga?
      Sira 1-fill-FV 7.cup   how in.what.sense
      'How did Sira fill the cup?'

a. geraha.	b. dududu.	c. * du.
slowly	DUDUDU	IDEO
'Slowly.'	'To the brim.'	

We propose that the triplication data in (42)-(43) involves the formation of (non-degree intensifying) adverbs. The semantics of the triplicated ideophones differs from the semantics of the non-triplicated ideophones in (34a) in that the triplicated ideophones (i) do not combine with a gradable predicate (i.e., something of type $<d<e,t>>$), and (ii) do not existentially quantify over degrees. We propose that the triplicated ideophones are similar to English "extreme" adjectives like *gigantic* and *gorgeous* in that they inherently pick out high degrees on their associated scale; in the case of *dududu*, the scale is one of fullness. These adverbs then freely distribute like other adverbs.

4 Conclusion

In this paper, we described the distribution and interpretation of ideophones in Llogoori, Lunyore, and Lutiriki. We sketched a preliminary proposal to treat Luhya ideophones as cross-categorial degree intensifiers, assuming the inclusion of degrees in the Luhya semantic ontology.

The Luhya ideophone data demonstrates the heterogeneity of ideophone systems cross-linguistically. Descriptively, the Luhya ideophones pattern very differently from highly depictive ideophone systems, which are often taken to be the norm. Theoretically, our degree-based proposal differs from other formal accounts of ideophones as depictions (Baglini 2016 for Wolof) and demonstrations (Henderson 2016 for Tseltal). Given the diversity of ideophone systems across languages, it is reasonable to postulate similar diversity in the formal theories used to account for them.

Finally, the data in this paper raises interesting questions with respect to the behavior of ideophones diachronically. Poulos (1999) posits that ideophones can undergo "grammaticalization,"

---

[20]The clefting data in (42) is only available in Llogoori; it is unavailable in Lutiriki and we do not have the relevant data for Lunyore.

which he associates with the loss of the ideophones' onomatopoeic properties and their eventual inclusion within an existing lexical category such as verbs or adverbs. We tentatively claim that the Luhya ideophones are in the process of being integrated into the set of Luhya adverbs. Some of the ideophones are less restrictive in their semantic class selection than others, suggesting that they are transitioning into being general degree intensifiers. For instance, *mno* picks out lexical items having to do with sweetness, happiness, and loneliness in Llogoori, whereas it distributes more freely in Lutiriki. Since Luhya speakers often live in highly multilingual environments, they may have begun borrowing ideophones from other languages (footnote 9); Mark Dingemanse (p.c.) notes that ideophones are among the first lexical items lost when speakers are no longer immersed in their language. However, much further diachronic study is needed to understand the trajectory of the Luhya ideophones.

References

Alpher, B. (1994). Yir-Yoront ideophones. In Hinton, L., Nichols, J., and Ohala, J., editors, *Sound Symbolism*, pages 161–177. Cambridge University Press.

Amha, A. (2001). Ideophones and compound verbs in Wolaitta. In Voeltz, F. and Kilian-Hatz, C., editors, *Ideophones*, volume 44 of *Typological Studies in Language*, pages 49–62. John Benjamins Publishing Company.

Antuñano, I. I. (2016). Basque ideophones from a typological perspective. University of Zaragoza, manuscript.

Baglini, R. (2016). The semantics of quotative ideophones. Handout, Semantics and Linguistic Theory 26.

Bartsch, R. and Vennemann, T. (1972). The grammar of relative adjectives and comparison. *Linguistische Berichte*, 21:168–185.

Bochnak, R. (2013). *Cross-linguistic Variation in the Semantics of Comparatives*. PhD thesis, University of Chicago.

Childs, T. (1988). The phonology of Kisi ideophones. *Journal of African Languages and Linguistics*, 10(2):165–190.

Childs, T. (1994). African ideophones. In Hinton, L., Nichols, J., and Ohala, J., editors, *Sound Symbolism*, pages 178–204. Cambridge University Press.

Cresswell, M. (1976). The semantics of degree. In Partee, B., editor, *Montague Grammar*, pages 261–292. Academic Press, New York.

Dhoorre, C. S. and Tosco, M. (1998). 111 Somali ideophones. *Journal of African Cultural Studies*, 11(2):125–156.

Diffloth, G. (1972). Notes on expressive meaning. In *Chicago Linguistic Society 8*, pages 440–447.

Dingemanse, M. (2011). *The Meaning and Use of Ideophones in Siwu*. PhD thesis, Radboud University Nijmegen, Nijmegen.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6(10):654–672.

Dingemanse, M. (2017). Expressiveness and system integration: On the typology of ideophones with special reference to Siwu. *STUF – Language typology and universals*, 70(2):119–141.

Dingemanse, M. and Akita, K. (2016). An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics*, pages 1–32.

Doke, C. (1935). *Bantu Linguistic Terminology*. Longmans, Green.

Dwyer, D. and Moshi, L. (2003). Primary and grammaticalized ideophones. In Mugane, J., editor, *Linguistic Typology and Representation of African Languages*, volume 5 of *Trends in African Linguistics*, pages 173–187. Africa World Press, Inc.

Essegbey, J. (2013). Touch ideophones in Nyagbo. In Orie, O. O. and Sanders, K., editors, *Selected Proceedings of the 43rd Annual Conference on African Linguistics*, pages 235–243, Somerville, MA. Cascadilla Proceedings Project.

Hamano, S. (1994). Palatalization in Japanese sound symbolism. In Hinton, L., Nichols, J., and Ohala, J., editors, *Sound Symbolism*, pages 148–157. Cambridge University Press.

Heim, I. (2001). Degree operators and scope. In Féry, C. and Sternefeld, W., editors, *Audiatur vox sapientiae*, pages 214–239. Akademie-Verlag, Berlin.

Henderson, R. (2016). Pluractional demonstrations. Handout, Semantics and Linguistic Theory 26.

Kennedy, C. (1999). *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. Garland, New York.

Kennedy, C. and Levin, B. (2008). Measure of change: The adjectival core of degree achievements. In McNally, L. and Kennedy, C., editors, *Adjectives and Adverbs: Syntax, Semantics, and Discourse*. Oxford University Press.

Kennedy, C. and McNally, L. (2005). Scale structure and semantic typology of gradable predicates. *Language*, 81(2):345–81.

Kubayi, S. (2009). The ideophone in Xitsonga: a morphosemantic analysis. Master's thesis, University of Limpopo.

Marivate, C. (1982). *The ideophone in Tsonga*. PhD thesis, University of South Africa, Pretoria.

Marlo, M. (2017). Tone in the Luyia varieties of western Kenya and eastern Uganda. Handout, 48th Annual Conference on African Linguistics.

Moltmann, F. (2009). Degree structure as trope structure: a trope-based analysis of positive and comparative adjectives. *Linguistics and Philosophy*, 32:51–94.

Morzycki, M. (2009). Degree modification of gradable nouns: size adjectives and adnominal degree morphemes. *Natural Language Semantics*, 17:175–203.

Msimang, C. and Poulos, G. (2001). The ideophone in Zulu: A re-examination of conceptual and descriptive notions. In Voeltz, F. and Kilian-Hatz, C., editors, *Ideophones*, volume 44 of *Typological Studies in Language*, pages 235–249. John Benjamins Publishing Company.

Newman, P. (1968). Ideophones from a syntactic point of view. *Journal of West African Languages*, 2:107–117.

Nicolas, D. (2004). The semantics of nouns derived from gradable adjectives. In Meier, C. and Weisgerber, M., editors, *Proceedings of Sinn und Bedeutung 8*, pages 197–2017.

Pérez González, J. (2012). Predicados expresivos e ideófonos en tseltal. Master's thesis, El Centro de Investigaciones y Estudios Superiores en Antropología Social.

Poulos, G. (1999). Grammaticalisation in South-Eastern Bantu and the linguistic 'dynamics' underlying this process. *South African Journal of African Languages*, 19(3):204–214.

Rett, J. (2008). *Degree Modification in Natural Language*. PhD thesis, Rutgers, The State University of New Jersey.

Samarin, W. J. (1971). Survey of Bantu Ideophones. *African Language Studies*, 12:130–168.

Schadeberg, T. (2003). Derivation. In Derek Nurse, G. P., editor, *The Bantu Languages*, pages 71–90. Routledge.

Simons, G. and Fennig, C., editors (2017). *Ethnologue: Languages of the World, Twentieth edition*. SIL International.

Voeltz, F. and Kilian-Hatz, C., editors (2001). *Ideophones*. John Benjamins Publishing Company.

Genericity in event semantics:
A look at Yoruba generic sentences[1]

Taofeeq Adebayo – Tulane University

Abstract. I propose a theory of genericity that is grounded in neo-Davidsonian event semantics (Parsons 1990, 2000; Higginbotham, 2000; etc.). I distinguish broadly between three types of individual: kind individuals, generic individuals and concrete individuals. A distinction is made between particular events and generic events on the one hand; and between kind-level states, individual-level states, stage-level states and generic states on the other hand. I propose that only generic individuals require the presence of the Gen operator and that kind and concrete individuals are existentially closed with the logical form of kind individual involving a type-shifting operation. Also, I propose that generic events and generic states contain the generic predicate '**gen** (e)' which turns concrete eventualities into generic ones and that the other types of eventuality also have their respective predicates that distinguish them from one another. When this framework is applied to genericity in Yoruba, it is shown first that Kimian states (Maienborn, 2007) in Yoruba have an E-position that the generic predicate (the imperfective) *máa-ń* targets (contrary to expectation), and second that in some constructions *máa-ń* is best treated as an overt realization of the operator Gen.

1   Introduction

My major concern in this paper is to propose a theory of genericity based on neo-Davidsonian event semantics (advanced in such works as Parsons (1990, 2000)), and to account for generic sentences in Yoruba. The main motivation for this proposal is based on the distinctions in (1).

(1)   a. Dogs bark.   [(generic *dogs*) (*bark* generically)]
        b. Dogs are friendly.   [(generic *dogs*) (are friendly)]
        c. John smokes a lot.   [(an individual *John*) (*smokes* generically)]
        d. It rains at night in Lagos.   (some event of raining occurs generically at night in Lagos)

According to Carlson (1989), generic sentences express regularities. Both entities (individuals) and eventualities (syntactically realized as VPs and their adjunctions) can be regular or generic in a generic sentence (where generic sentences are taken to include habituals and reference to kind). Generic entities and eventualities can occur together as in (1a) and independently as in (1b-d).

Assuming that the distinctions made in (1) are accurate, this paper argues for two sources of genericity in the syntax which can both be combined in a single clause as well as function independently as shown in (1). I argue that a proper understanding of genericity requires a distinction between generic, kind and concrete individuals and between generic and non-generic eventualities. As a result, we can distinguish between entity-driven genericity and eventuality-

---

driven genericity. This distinction is close to Krifka et al's. (1995) distinction between characterizing statements and reference to kind but differs in a number of ways that become clear below. Languages differ in what sorts of NPs and predicates can be generic based on the configuration of the syntax. However, there are a number of cross-linguistic generalizations that can be made on the semantics of generic individuals and eventualities. For example, we can make a pre-theoretical assumption that all languages have generic individuals and eventualities as well as other kinds of individuals and eventualities, which are considered below.

The rest of this paper is organized as follows. I lay out a theoretical framework for genericity in event semantics in Sections 2 and 3. I then apply this to Yoruba generic sentences in Sections 4 and 5. I identify some preliminary advantages of the framework in section 6. The two most important conclusions in this section are that Kimian states in Yoruba do have an E-position and that the Operator Gen is pronounced in some constructions in the language. Section 7 concludes.

2   Three ontological types of individuals

The many theories of NP/DP referencing in generic sentences include: uniform kind-referencing (Carlson, 1977 and others), neo-Carlsonian Approach (Chierchia, 1998 and others), the Ambiguity Hypothesis (Gerstner-Link and Krifka, 1993 and others), theory of incorporation (Farkas and de Swart, 2003 and others). For an elaborate description of the development of theories on this topic, see Mari et al. (2013). But works on NP/DP referencing that are directly relevant to the ontological distinction made below are those that address the distinction between specific and nonspecific NPs. Discussions of this distinction can be found in Baker (1973), Hawkins (1978) Fodor and Sag (1982) and Enç (1991). A common distinction between the two relates to the scope of the NP. NPs with narrow scope are generally taken to be nonspecific (*a student* in (2c)) while those with wider scope are taken to be specific (*a student* in (2b)).

(2)   a. Every teacher in that school beat a student yesterday.
        b. **A student** was such that all the teachers in that school beat him yesterday.
        c. For each of the teachers in that school, there was **a student** they beat yesterday.

Another distinction that plays a major role in the ontology developed below can be found in Krifka et al. (1995). According to Mari et al. (2013:6), Carlson (i.e. 1977) was the first to propose an ontology for kind, distinguishing it from 'normal individuals' like John. This ontology is recognized and advanced in Krifka et al. (1995). The NP 'dinosaurs' in 'Dinosaurs are extinct' refers to the kind individual, 'dinosaur kind'. Krifka et al. (1995) also make a distinction between specific and non-specific NPs but this distinction was different from the traditional distinction that is based on the difference between (2b) and (2c). For them, a specific NP refers to a particular individual while a nonspecific NP does not refer to any specific individual in particular. In the following examples (Krifka et al. 1995:16), 'a dog' (3a) refers to a specific dog while 'a dog' (3b) refers to dogs in general.

(3)   a. **A dog** is barking.   b. **A dog** barks.

The result of the above discussion is that there is a conflict of terminology regarding what is a specific NP and what is not. Recognizing this conflict, Krifka et al. (1995:15) put their own classification on a purely pre-theoretical level. In what follows, I attempt to resolve this conflict in the categorization that I propose. Now let us consider what we have established up to this point. First, we have established, based on Carlson (1977), that there is such a thing as a kind individual distinct from ordinary individual. Making this distinction also makes us realize the existence of normal or ordinary individual, which is often referred to in the literature as specific individual (e.g. Krifka et al., 1995:15 and Pelletier, 2010:11). Two kinds of individuals are thus sufficiently recognized in the literature: kind individuals (e.g. *dinosaurs* in the example above) and specific individuals (*a dog* in (3a)). A general distinction that Carlson (1977:442) makes between kind individual (or kind-level individual) and specific (or normal) individual is that the former can be here and there while the latter is confined to a location at a given time. Using this diagnostic, we can posit that the individuals referred to in (2b), (2c) and (3a) are normal individuals (in the sense that we have been using the term 'normal'), while *dinosaur* refers to kind individual. However, the reference that the NP *dogs* in (3b) make is not as quite determinable using the kind-normal distinction above. It does not refer to a dog located at a particular location and time, and it does not refer to a kind the same way that *dinosaurs* does. This, therefore, forces us to recognize another kind of individual that is in a medial position between kind individual and normal individual: an individual which is not specific and is abstract like kind individual, but which is different from kind individual in that it accumulates its properties from generalizations about instances of a kind. I call this 'generic individual' and assume that it is this abstract individual that NPs like *a dog* in (3b) refer to. The idea of a generic individual is not unheard of. A similar idea can be found in the philosophy literature. Fine (1983), for instance, argued in defense of a long-standing idea about the concept of arbitrary objects which are distinguished from individual objects.

Let us now go back to the conflict of terminology identified above. In all of the examples above where we have identified specific or normal individuals, the common characteristic of all the NPs is that they refer to concrete instances which are located in time and space. For this reason and to escape the terminological problem associated with the specific-non-specific distinction, I will refer to this kind of individual as 'concrete individual' which can, then, be subdivided into different categories to account for the differences among (2b), (2c) and (3a). Consider the following examples of the three kinds of ontological individuals established so far:

(4)   a. **Dogs** are everywhere. [kind indiv.]   b. **A dog** barks. [generic indiv.]
        c. **A dog** is barking. [concrete indiv.]   d. Everyone brought **a dog** each.[concrete indiv.]
        e. **The dog** is barking. [concrete indiv.] f. **Jack** barks at night. [concrete indiv.]

We can now examine each of these individuals one at a time. I start with concrete individuals. These individuals have nuances that make them distinct from one another. The difference we pointed out between (2b), (2c) and (3b) suggests strongly that we must be able to distinguish different kinds of concrete individuals in our system. At this point, we can recognize at least four types: specific, non-specific, definite and proper. This is illustrated as follows:

(5)   a. **A dog** is barking.   [concrete **specific** individual]
        b. Everyone brought **a dog** each.   [concrete **non-specific** individual]

c. **The dog** is barking.   [concrete **definite** individual]
d. **Jack** barks at night.   [concrete **proper** individual]

I assume that these types have the same general logical form which can be modified variously to account for the minor differences. For example, (5a), will have the following logical form, leaving out events for now: $\exists x\ [dog(x) \wedge barks(x)]$. We can appeal to scope position to distinguish (5b) from (5a). For (5c), we can make use of the iota notation. For (5d), we might consider 'Jack' a referring expression (as in $[barks(j)]$) or a predicate bound by existential closure as in (5a-c). What seems to be common to each is binding by existential closure.

Next, I consider the kind individual. According to the general assumptions of Carlson (1977), kind individuals are abstract individuals that may have actual instantiations. Since they are not concrete, they lack the kind of regular existential closure used for concrete individuals. There are broadly two ways in which we can implement the logical form of kind individuals. We can treat them as proper names with direct kind predication as in (6b); we can also assume that they involve variables with existential closure which are type-shifted into the kind individual. The latter way can be implemented in the various ways shown below (6c-e).

(6)   a. Dinosaurs are extinct.   b. extinct (dinosaurs)
        c. $\exists x\ [dinosaur_k\ (x) \wedge extinct\ (x)]$ (after Krifka et al 1995)
        d. $\exists x\ [\uparrow dinosaur\ (x) \wedge extinct\ (x)]$ (after Link 1995:382)
        e. $\exists x\ [^{\cup \cap}dinosaur\ (x) \wedge extinct\ (x)]$ (after Chierchia 1998)

Another way to implement (6) is to assume that there is a Gen operator, which binds a type-shifted variable that refers to the kind. I assume here that kind individual, as well as concrete individual, does not need the Gen operator and that it is only the generic individual that the Gen operator binds. This assumption is based on the following argument. I assume here that an inherent part of the concrete individual and the kind individual is that they are existentially identified in terms of ontology. For example, when we say *a dog is outside*, we can paraphrase this as 'there is something in the world namely a dog that is outside'; likewise, when we say *dogs are widespread*, we can paraphrase this as 'there is some kind in the world namely the kind dog that is widespread'; it will be awkward, however, to paraphrase a statement like *dogs bark* as 'there are some individuals in the world namely some dogs that bark'. Generic individuals are not existentially closed at least not in the sense that concrete and kind individuals are.

One of the major assumptions of this paper is that there is a third kind of individual that is neither concrete nor is a kind individual. I identified this as the generic individual. Like kind individual, it is an abstract individual, but it gets its own properties from generalizations about instances of a kind. I assume here that since they are based on rough generalizations about instances of a kind individual, they are bound by the unpronounced Gen operator as in the following: Dogs bark = $Gen\ x\ [dog(x) \rightarrow bark\ (x)]$.

3   Eventualities

Davidsonian events semantics (Davidson, 1967) assumes that verbs of action and change in natural language have a hidden event argument now commonly referred to as the E-position (see

Higginbotham, 1985:555). This idea has been extended in the neo-Davidsonian tradition (represented in such work as Parsons 1990 and 2000; Landman, 2000; Higginbotham, 2000, etc.) which assumes that all predicates, including statives, have the E-position. While this has been widely accepted in the literature, scholars such as Maienborne (e.g. Maienborn, 2007 and 2011) and Katz (e.g. Katz, 2000) have continued to argue for a Davidsonian view that only eventive predicates (Katz, 2000) and 'Davidonian states' like *sit* and *sleep* (Maienborn, 2007) have an E-position. Maienborn (2007), for example, demonstrates that a kind of state she describes as Kimian state, ontologically and linguistically, defers from eventive predicates and Davidsonian states in a number of ways, arguing that they lack the E-position.

For reasons of space, I do not address this distinction between eventives and Kimian states or what Moltmann (2013) calls 'abstract states' here but see Section 6 for the consequences of the proposal here for Kimian states. In the theoretical framework that I lay out shortly, I assume that all predicates have an E-position. I also assume (following Parsons, 1990) that there are three sorts of things that predicates generally encode: events, states, and processes, which I collectively refer to as 'eventualities' (Bach, 1986). For the purpose of this paper, my focus is on events and states. First, I propose that there are different kinds of states and events with regard to their interaction with individuals, their duration and number of instances across time and space. For example, states such *is extinct*, only apply to kind individuals, a state such as *is hungry* is shorter in duration than a state like *is intelligent*, while a particular event such as *smoked last night* has one instance whereas an event such as *smokes after dinner* has multiple instances. It is based on these facts that the following ontological distinctions are made. I start with events.

3.1   Events

The general tradition in event semantics is to think of events in concrete terms. Events are located in space and time. We can have multiple events such as e_1, e_2, etc.; we can also have subevent (e') commonly proposed for such constructions as resultatives and causative-inchoatives. What is not common is to think of events as having uncountable instances. But there are some events which cannot be given the description of a particular event and which seem to have multiple instances that are not countable. Consider the following:

(7)   a. Mary smoked at the party.   b. ∃e [**smoking** (e) ∧ **ag** (e, Mary) ∧ **at_the_party** (e)]

(8)   a. Mary smokes after dinner.   b. * ∃e [**smoking** (e) ∧ **ag** (e, Mary) ∧ **after_diner** (e)]

While the interpretation in (7b) is accurate for (7a), ignoring tense and salience, (8b) does not give an accurate interpretation of (8a) for the reason that (8b) suggests that there is a concrete event which took place at a certain time but which does not take place with some regularity. This is against the meaning of (8a). Although (8a) can be given the standard analysis in the following way: GEN [x,s;] (x = Mary & smoke (x. s); after.dinner (s)) (Krifka, 1995:238), the question is: can event semantics handle the regularity that is associated with some events such as the one in (8a)? To answer this question, I first propose that there are two types of events. These are particular/concrete events and generic events. This particular distinction can be found in the work of Montague (Montague, 1969) as reported in Pianesi and Achille (2013). According to

them, Montague distinguished between generic events (e.g. *sun rises*) and particular events (e.g. *sun rose yesterday*). For him, generic events are a kind of property and particular events are instantiations of generic events. Montague's theory is metaphysical however, and its major assumption is that particular events are derived from generic events. In the system that I advocate in this paper, generic events are derived from particular events with an addition of a predicate. My own distinction between generic and particular/concrete events, therefore, is purely linguistic and ontological rather than broadly metaphysical. Let us start by observing the following:

(9)   There are ontologically two kinds of events in natural language:
        a. Particular/concrete events with countable instances
        b. Generic events with uncountable instances

I define particular events as a kind of event that is located in a specific time and location and does not express any form of regularity. I also assume that they have single instances. Of course, they can be distributive as in *John buttered three loafs of bread*, and have subevents, but they generally lack the property of having uncountable instances. A generic event, on the other hand, is an event which expresses some regularity. A generic event is true only when there is some regularity involved and it is not the case that it has definite instances. For instance, in the event in the expression *Africans drink palm wine*, it is hard to think of how many instances this event *drink* has. To formalize this regularity in the neo-Davidsonian framework, I propose that all events have an additional predicate that is either concrete (10a) or generic (10b).

(10)   a. [[concrete]] $_{\langle v,et\rangle}$ = λPλe [P (e) ∧ **con** (e)]   b. [[generic]] $_{\langle v,et\rangle}$ = λPλe [P (e) ∧ **gen** (e)]

When (10b) is applied to a standard neo-Davidsonian event, it turns such an event into one that occurs with some regularity and allows for counterfactuals, whereas when (10a) is applied to an event, it indicates that the event has a concrete instance. The distinction between (7a) and (8a) can now be handled as in (11a&b) respectively.

(11)   a. [λPλe [P (e) ∧ **con** (e)]] (∃e [**smoke** (e) ∧ **ag** (e, **mary**) ∧ **at_the_party** (e)]) =
            ∃e [**smoke** (e) ∧ **con** (e) ∧ **ag** (e, **mary**) ∧ **at_the_party** (e)]
        b. [λPλe [P (e) ∧ **gen** (e)]] (∃e [**smoke** (e) ∧ **ag** (e, **mary**) ∧ **after_dinner** (e)]) =
            ∃e [**smoke** (e) ∧ **gen** (e) ∧ **ag** (e, **mary**) ∧ **after_dinner** (e)]

This derivation defers from the standard treatment of events only with the introduction of the generic/concrete predicate. Introducing a new predicate is not uncommon in the event semantics literature. Parson (1990:28) for example, uses **Cul** and **Hold** to account for tense and aspect; **gen/con** is no less a predicate accounting for a functional category. Summarily, what distinguishes concrete events from generic events is the generic/concrete predicate.

3.2   States

Four distinct types of states can be identified in terms of duration, number of instances and the kind of individual that can be their argument. Let us start with (12):

(12)   Four ontological kinds of states can be identified:
       a. kind-level state   b. stage-level state
       c. individual-level state   d. generic state

Kind-level states are a kind of state that only takes a kind individual as a theme. For example, *is extinct* is a state that can only be true of the kind *the dinosaur*. I borrow Carlson's (1977) terminology of individual-stage-level predicates and distinguish between stage-level states and individual-level states. A stage-level state is one which applies to stages of an individual; this is tantamount to what one might regard as temporary state. *Is hungry* is an example of stage-level states. An individual-level state applies to each and every stage of an individual. This is what one might consider a permanent state. An example of this is *is brave*. Generic state is a kind of state that comes with some regularity and has multiple instances that are not definite. An example of this is *is always hungry*. This is neither a pure stage-level state nor an individual-level state. It is a state that is scattered among the stages of an individual and does not apply to each and every stage of such individual. I assume that the mechanism in (10) is applicable to the categories in (12) and use the predicate **kind-l** (s) for kind-level states, **stage-l** (s) for stage-level states and **ind-l** (s) for individual-level states. But for the sake of space, these predicates, and the '**con** (e)' predicate are not indicated in the notations in subsequent sections. The predicate '**gen** (s)' for generic state and '**gen** (e)', however, are indicated, since they are the focus of the paper. The following examples illustrate the four types of state identified above:

(13)   a. Dinosaurs are extinct
       b. ∃s [**being_extinct** (s) ∧ **kind-l** (s) ∧ **th** (s, ↑dinosaurs)]   kind-level state

(14)   a. John is hungry.
       b. ∃s [**being_hungry** (s) ∧ **stage-l** (s) ∧ **th** (s, john)]   stage-level

(15)   a. John is clever.
       b. ∃s [**being_clever** (s) ∧ **ind-l** (s) ∧ **th** (s, john)]   individual-level

(16)   a. John is always hungry.
       b. ∃s [**being_hungry** (s) ∧ **gen** (s) ∧ **th** (s, john)]   generic

The reader might find it unusual that the adverb of quantification *always* in (16a) has been reduced to a predicate in (16b), given that it is often treated as an operator. In the next section, I show that Yoruba treats both generic state and event the same, so that (16a) is expressed with the same grammatical means that generic events are expressed with.

4   Generic eventualities in Yoruba

In this section, I provide an account of how the various types of eventualities I have identified above are realized in Yoruba, with a focus on generic events and states. Let us start with non-generic eventualities. These eventualities have the common characteristic that they are realized in the syntax with simple predication. No additional particle or marker is needed to express them.

Generic events and states, on the other hand, are explicitly marked in the language. A generic eventuality is generally marked with the imperfective marker *máa-ń*. This marker has the function of taking particular events and turning them to events with indefinite instances. This is exactly what the **gen** predicate proposed above does. Let us give the logical form of (17a) as in (17b) and the denotation of *máa-ń* as in (18). (19& 20) show application of (18) to (17b).

(17) a. Bólá jẹ ewé
  Bólá eat leaf
  'Bólá ate leaves.'
  b. ∃e [**eating** (e) ∧ **ag** (e, bólá) ∧ **th** (e, leaf)]

(18) [[máa-ń]] = [[generic]] ⟨v,et⟩ = [λPλe [P (e) ∧ **gen** (e)]]

(19) Bólá **máa-ń** jẹ ewé ('Bólá eats leaves (habitually'))

(20) a. [λPλe [P (e) ∧ **gen** (e)]] (∃e [**eating** (e) ∧ **ag** (e, b) ∧ **th** (e, leaf)])
  b. ∃e [**eating** (e) ∧ **gen** (e) ∧ **ag** (e, **bólá**) ∧ **th** (e, **leaf**)]

While (17b) states that there is a particular one-instance event of eating leaves that has Bólá as an agent, (20b) states that there is an indefinite multiple-instance event of eating leaves that has Bólá as an agent. The implication of (18), therefore, is that the category of events that was identified as generic event in the previous section not only has an ontological support but also a linguistic support in Yoruba. Next, let us consider the case of generic state.

Linguistic support for the category of generic state is not readily available in English since what corresponds to *máa-ń* is not phonologically available. But this support is found in Yoruba. Generic states, just like generic events, are constructed from particular states (stage-level states in most cases) by using *máa-ń* (the generic predicate). Consider the following:

(21) a. Bólá wà ní ilé ọtí
  Bólá exist in house alcohol
  'Bólá is/was at the bar.'
  b. Bólá **máa-ń** wà ní ilé ọtí
  Bólá **gen** exist in house alcohol
  'Bólá is/was at the bar in multiple indefinite instances (habitually).'
  c. ∃s [**being-in-the-bar** (s) ∧ **gen** (s) ∧ **th** (s, bólá)]

(21c) which is the logical form of (21b) states that there is a generic state of being at the bar whose theme is *bólá*. Note that the generic state in (21b) can also be expressed in English as 'Bólá is always at the bar'. But this cannot give the accurate information that is expressed in this sentence, because (21b) does not contain anything that corresponds to adverb of quantification. It only states that there are multiple occasions of Bólá being at the bar, and does not specify whether this is usually, seldom or always. To do that, prepositional constructions that are similar to English adverb of quantification (e.g. *ní ẹ̀ẹ̀kòọ̀kan* 'sometimes/seldom', *ní ọ̀pọ̀ ìgbà* 'often/usually/ in most cases', etc.) will have to be used. The consequent intuition, therefore, is

that *máa-ń* is a true generic predicate that modifies an eventuality variable to give it the property of having multiple instances of unspecified number.[2]

5 Concrete, generic and kind individuals in Yoruba

There are only two forms of NP in Yoruba that make reference to kind individuals and generic individuals. These are what I refer to in this paper as Bare NPs (BNPs) like *ajá* 'dog', *ewúrẹ́* 'goat', etc., and Plural NPs (PNPs) such as *àwọn ajá* 'dogs', *àwọn ewúrẹ́* 'goats', etc. These two forms can also make reference to the different kinds of concrete individuals identified in Section 2. The starting point then is to assume that these two forms are ambiguous between reference to kind, generic and concrete individuals. What determines which individuals they refer to is the type of eventuality in which they are serving as argument and the nature of the second argument in the eventuality. Let us start with the kind individual. The two forms are interpreted as kind individuals when they serve as a theme of a kind-level state (henceforth, K-state) as shown in (22ai). But only the BNP yields kind interpretation with individual-level sate (henceforth, I-state) as in (22bi); PNP is odd in this context. Also, when both forms are a theme of an I-state with an experiencer (see footnote 2), kind interpretation is obtained as in (22ci). Generic interpretation of BNP is also obtained in deontic modality even when the eventuality is a particular event as shown in (22di). In this latter case, it is assumed that the deontic modality turns a concrete event into a state (property) which can be predicated of a kind as shown in (22dii). This property can then be inherited by members of the kind in appropriate worlds.

(22)   a. BNP/PNP as theme of K-state → kind individual
            i. (Awọn) Ajá   wà   káàkiri
               (PL)   dog   be   everywhere
               'Dogs are everywhere.'
           ii. ∃s [**being-everywhere** (s) ∧ **th** (s, ↑dog)]
        b. BNP as theme of I-state → kind individual
            i. Ewúrẹ́   ní   ẹ̀jẹ̀
               goat   have   blood
               'Goats have blood.'
           ii. ∃s [**having-blood** (s) ∧ **th** (s, ↑goat)]

---

c. BNP/PNP as theme of K-state with an experiencer[3] → kind individual
  i. Bọ́lá   fẹ́ràn   (àwọn) ọmọdé
    Bọ́lá   like   (PL)   child
    'Bọ́lá loves children.'
  ii. ∃s [**loving** (s) ∧ **exp** (s, bọ́lá) ∧ **th** (s, ↑child)]
d. BNP as an agent of particular event in deontic modality → kind individual
  i. Ayékòótọ́   lè   kọrin
    parrot   can   sing
    'A parrot can sing'
  ii. λw ∃e [singing (e) ∧ ∃s [able-to-be-agent-of-'e' (s) ∧ th (s, ↑parrot) ∧ in (s, w)]]

Let us next consider the concrete individual. The two forms are interpreted as concrete individuals when they are an agent or a theme of a particular event (henceforth, P-event) as in (23ai). They are interpreted as concrete individuals when they are an agent (23bi) or a theme (23ci) of a generic event (henceforth, G-event) with a concrete individual argument.

(23)   a. BNP/PNP as an agent or theme of P-event → concrete individual
      i. (Àwọn) ajá   jẹ   egungun
        (PL)   dog   eat   bone
        'A/the dog/ the dogs ate a piece (some pieces) of bone.'
      ii. ∃e [**eating** (e) ∧ **ag** (e, **dog**) ∧ **th** (e, **bone**)]
    b. BNP/PNP as an agent of G-event with concrete individual theme→ concrete individual
      i. (Àwọn) ajá   maa-n lé   Bọ́lá
        (PL)   dog   **gen** chase Bọ́lá
        'A certain dog or some certain dogs chase Bọ́lá in indefinite occasions'
      ii. ∃e [**chasing** (e) ∧ **gen** (e) ∧ **ag** (e, **dog**) ∧ **th** (e, **bọ́lá**)]
    c. BNP/PNP as a theme of G-event with concrete individual agent → concrete individual
      i. Bọ́lá   maa-n lé   (àwọn) ajá
        Bọ́lá   **gen** chase (PL)   dog
        'Bọ́lá chases a certain dog or some certain dogs in indefinite occasions'
      ii. ∃e [**chasing** (e) ∧ **gen** (e) ∧ **ag** (e, **bọ́lá**) ∧ **th** (e, **dog**)]

Concrete individual interpretation is also obtained when BNPs and PNPs serve as the theme of a stage-level state (henceforth, S-state) (24ai) and as the experiencer of an I-state with concrete individual theme (24bi). They are also interpreted as concrete individuals when they are an experiencer in a generic state (henceforth, G-state) that has a concrete individual argument (24ci) or when they are a theme of a G-state with a concrete individual experiencer (24di).

(24)   a. BNP/PNP as a theme of S-state → concrete individual
      i. (Àwọn) ajá   dákẹ́
        (PL)   dog   be.silent

---

[3] I use the term 'experiencer' to refer to the individual who is aware of a stimulus (following Hilpert, 2014:27). The term is used here to distinguish between the arguments of verbs like *love*. For example, in *Jane loves dogs*, *Jane* is the experiencer and *dogs* is the theme. I use the notation 'exp' to signify the term 'experiencer' in logical forms.

'The dog(s) are silent or the dog(s) became silent.'

    ii. ∃s [**being-silent** (s) ∧ **th** (s, **dog**)]

b. BNP/PNP as an experiencer of S-state with concrete individual theme → concrete individual

    i. (Àwọn) ajá   fẹ́ràn   mi

    (PL)   dog   like   1SG

    'The dog(s) like me.'

    ii. ∃s [**liking** (s) ∧ **exp** (s, dog) ∧ **th** (s, **me**)]

c. BNP/PNP as an experiencer of G-state with concrete individual theme → concrete individual

    i. (Àwọn) ajá   máa-ń fẹ́ràn   mi

    (PL)   dog   **gen**   like   1SG

    'A dog/ some dogs like me in indefinite number of occasions'

    ii. ∃s [**liking** (s) ∧ **gen** (s) ∧ **exp** (s, **dog**) ∧ **th** (s, **me**)]

d. BNP/PNP as a theme of G-state with concrete individual experiencer → concrete individual

    i. Mo   máa-ń fẹ́ràn (àwọn) ajá   pupa

    1SG   **gen**   like   (PL)   dog   red

    'I like red dog(s) in indefinite number of occasions'

    ii. ∃s [**liking** (s) ∧ **gen** (s) ∧ **exp** (s, **I**) ∧ **th** (s, **dog**)]

A generalization that can be observed with regard to the interpretation of BNPs and PNPs as concrete individuals is that they require an aspect of a proposition to be concrete or to be located in time and/or place. That is, they require that either the eventuality is particular or that there be a second argument that is particular (concrete). For instance, we see in (23) and (24), that concrete-individual interpretation is tied to P-events (23ai), S-state (24ai), and the requirement that the second argument has a concrete-individual interpretation (23bi, 23ci, 24bi, 24ci and 24di). It should be noted as well that there are different types of concrete individual in (23) and (24). For instance, *ajá* 'dog' and *egungun* bone in (23ai) refer to concrete specific individuals, *ajá* 'dog' and its plural form *àwọn ajá* 'dogs' in (23bi, 23ci, 24ci and 24di) refer to non-specific individual, *ajá* 'dog' and *àwọn ajá* 'dogs' in (24ai) and (24bi) refer to concrete definite individual, while *Bọ́lá* in (23bi) refers to a concrete proper individual.

Let us start by observing that, generally, only BNPs are naturally interpreted as generic individual; PNPs either yield existential interpretation or are generally odd. One peculiar characteristic of interpreting BNPs as generic individuals is that they occur in generic eventualities. However, it should be noted that things are not as quite straightforward with this observation, as there are some of these eventualities that superficially appear as generic but are best analyzed as non-generic. The starting point then is to make the distinction between true generic eventualities in this regard and superficial generic eventualities and then see how the BNPs figure. Examples of true generic eventualities that yield generic interpretation of BNPs is given in (25). (25ai) shows how BNPs are interpreted as generic in G-event with no theme while (25bi) demonstrates generic intepretation in G-event with a concrete non-specific theme.

(25)   a. BNP as an agent of G-event with no theme → generic individual
      i. Ajá   máa-ń gbó
      dog   **gen**   bark
      'Dogs bark.'
      ii. Gen x [**dog** (x) → ∃e [**barking** (e) ∧ **gen** (e) ∧ **ag** (e, x)]]

    b. BNP as an agent of G-event with concrete non-specific theme → generic individual
      i. Ajá   máa-ń jẹ   egungun
      dog   **gen**   eat   bone
      'Dogs eat bones.'
      ii. Gen x [**dog** (x) → ∃e [**eating** (e) ∧ **gen** (e) ∧ **ag** (e, x) ∧ **th** (e, bone)]]

What (25) basically shows is that when BNPs occur in generic eventualities, they are interpreted generically. The source of their generic interpretation can then be located in those generic eventualities. However, note that (25aii) and (25bii) defers from the standard Gen approach in two respects: first, the verbal predicate is interpreted as event and this event is taken to be generic; second, the generic interpretation of the NP is said to be tied to the genericity of the eventuality. As such, (25aii) states that generally for dogs there is some generic event of barking that they do, while (25bii) states that generally for dogs there is some generic event of eating concrete (non-specific individual) bones that they do. Next, consider the superficial generic eventualities that also yield generic interpretation for BNPs:

(26)   a. BNP as an experiencer of superficial G-state with kind theme → generic individual
      i. Ajá   máa-ń fẹ́ràn egungun
      dog   **gen**   like   bone
      'Dogs likes bone.'
    b. BNP as an experiencer of superficial G-state with concrete non-specific theme →
      generic individual
      i. Ewúrẹ́ máa-ń ní   ìwo
      goat   **gen**   have horn
      'Goats have horn'
    c. BNP as a theme of superficial G-state → generic individual
      i. Mààlû   máa-ń tóbi
      cow   **gen**   be.big
      'Cows are big.'

We can go ahead and give (26ai, bi and ci) the same kind of treatment as before, so that their logical forms are as in (27a-c) respectively. But this will be counter-intuitive as argued below.

(27)   a. Gen x [**dog** (x) → ∃s [**liking** (s) ∧ **gen** (s) ∧ **exp** (s, x) ∧ **th** (s, **bone**)]]
    b. Gen x [**goat** (x) → ∃s [**having** (s) ∧ **gen** (s) ∧ **exp** (s, x) ∧ **th** (s, **horn**)]]
    c. Gen x [**cow** (x) → ∃s [**being-big** (s) ∧ **gen** (s) ∧ **th** (s, x)]]

(27) states, wrongly, that some generic individuals undergo an I-state in an indefinite number of occasion. For instance, (27c) states that generally for cows, there is some state of being big

that they experience in an indefinite number of occasions. This is contrary to the meaning of (26ci) which only says that there is an indefinite number of occasions where a given cow is big. To resolve this mismatch, we have to do away with the generic predicate in (27) and treat the eventualities therein as I-states rather than G-states constructed from I-states. The implication of this then is that the source of generic interpretation for the BNPs in the examples in (26) cannot be located within the eventualities but must be from a different source in those sentences.

The most available intuition is that, if the generic interpretation of the BNPs cannot be due to the eventualities in those sentences, then it must be due to the imperfective marker *máa-ń*, which has been argued above to be the generic predicate. The examples in (26) are different. *Máa-ń* does not turn the eventualities in those sentences to generic eventualities, but instead ensures that the BNPs in those sentences are interpreted generically. If this observation is in the right direction, then *máa-ń* does not serve as the generic predicate in those sentences, but as an operator that binds the variables supplied by the BNPs. From this viewpoint, *máa-ń* in (26), therefore, has a semantics that is very close to or the same as that given to the generic operator, Gen. We can then posit that *máa-ń* has the two denotations shown below:

(28)   a. [[máa-ń]] = [[generic]] $_{\langle v, et \rangle}$ = [λPλe [P (e) ∧ **gen** (e)]]
        b. [[máa-ń]] = [[Gen]] $_{\langle e, t \rangle}$ = λPλQ Gen x [P (x) → ∃e/s [Q(e/s) ∧ **ag/th** (e/s, x)]]

Let us illustrate (28b) with the example in (26bi) given as (29b) below. Recall that BNPs that serve as a theme/experiencer of an I-state are interpreted as kind individuals (see 22b). Without the imperfective marker *máa-ń*, the BNP in (26bi) is interpreted as a kind individual as in (29a). Consider the following:

(29)   a. Ewúrẹ́ ní ìwo = ∃s [**having-horn** (s) ∧ **exp** (s, ↑goat)] = kind individual
        b. Ewúrẹ́ máa-ń ní ìwo =
        c. Máa-ń$_x$ [Ewúrẹ́$_x$ ní ìwo] =
        d. Gen$_x$ [Ewúrẹ́$_x$ ní ìwo] =
        e. Gen x [**ewúrẹ́** (x) → ∃s [**níní** (s) ∧ **exp** (s, x) ∧ **th** (s, ìwo)]] =
        f. Gen x [**goat** (x) → ∃s [**having** (s) ∧ **exp** (s, x) ∧ **th** (s, **horn**)]]

(29b-f) demonstrates the compositional derivation for *máa-ń* as an operator. (29c) demonstrates that *máa-ń* specifically targets a variable that is supplied by the BNP. The other examples in (26) has to be given the same logical form in (29f) where there is no generic predicate, but rather a generic operator that is phonologically available.

If (28b) is correct and the derivation for *máa-ń* in (29b-f) is accurate, then it follows that the so-called silent operator Gen, may not be silent in some languages and some contexts after all. The discussion above has shown that, while it is silent in some constructions as in the examples in (25) where its presence is due to the nature of the eventuality, it has a pronounced counterpart in other constructions, as demonstrated by the examples in (26). The implication, therefore, is that Yoruba provides an empirical support for the so-called Gen operator.

A generalization that can be taken from the discussion so far is that BNPs are naturally interpreted as abstract individual (kind or generic) while PNPs naturally have an existential interpretation of concrete individuals. It was shown that PNPs are possible as kind individuals

but this is rather far restricted. The fact that PNPs are generally odd as generic individuals also suggest that PNPs are naturally existential and that their interpretation as kind individual is rather due to a type-shifting operation whose source can be located in the eventuality. Existential interpretation of BNPs can also be explained away by a type-shifting operation that is occasioned by the eventuality. This type-shifting operation might be reminiscent of Carson's theory, but I am not committed to that theory. I have only employed this conceptualization ad hoc to put the general distribution of these two forms of Yoruba NPs in proper perspective.

6   Some preliminary advantages

The most important advantage of the theoretical framework proposed in this paper is the freedom it allows for one to look at genericity in a language like Yoruba that has not been robustly researched in this area. I was able to account for genericity in Yoruba without having to commit to any specific theory of genericity whose limitations could have hindered exploratory pursuits. The framework also has cross-linguistic applicability. For languages whose generic sentences have not been researched before, the framework provides a general adaptable analytic guideline that makes use of ontological distinctions in individuals and eventualities which may be distinguished in different ways cross-linguistically.

The framework is also able to account for not only subject arguments but also object arguments in terms of generic-kind-existential interpretation. Most theories of genericity concentrate on subject arguments (see Mari et al., 2013:2), but it seems that we need to be able to account for object arguments as well. The ability of this framework to account for both subject and object arguments means we can account for grammatically conditioned genericity in Yoruba as found in the syntactic account of Ajiboye (2005). As we saw above, sometimes, generic-kind-existential interpretation of an NP may be conditioned by the nature of the second argument. Using events semantics makes it possible to account for this fact.

This framework also avoids recourse to pragmatics, as is often the case in the standard Gen approach. For instance, (1d) repeated in (30a) will be given a logical form like that in (30b). According to this framework, the logical form of (30a) can be restated as in (30c).

(30)   a. It rains at night in Lagos.
        b. Gen s [s is a situation appropriate for raining in Lagos → it rains]
        c. $\exists$e [**raining** (e) $\land$ **gen** (e) $\land$ **at** (e, **night**) $\land$ **in** (e, **lagos**)]

(30c) does not specify more than what is present compositionally in (30a). Since there is no generic individual in (30a), there is no need for the Gen operator. This is consistent with the assumption of the framework that only generic individuals are bound by the Gen operator.

Another significance of this framework is in the fact that it makes some predictions that may have bearing on current issues in event semantics. Maienborn's theory of statives and copula + adjectives (Maienborn, 2004, 2005 and 2007) states that statives and copula + adjective are Kimian states that lack Davidsonian event argument. If this assumption were true for Yoruba, then there would be nothing we can refer to as generic state. But as the discussion above has shown there is indeed a generic state in the language. Let us consider this here again:

(31)   a. Ade   bínú   b. Adé máa-ń bínú
       Ade   be.angry   Ade **gen**   be.angry
       'Ade was angry.'   'Ade is *generically* angry.'

First, note that (31b) is not available in English: the approximation 'Ade is usually angry' does not accurately represent the meaning of sentences like (31b) in that it does not contain any prepositional phrase that are equivalent to English adverb of quantification. (31b) only states that Ade is angry in multiple number of occasions that is not definite just as 'Mary smokes' indicates that Mary smokes in an indefinite number of occasions. According to Maienborn's theory, (31b) should not be possible since *bínú* 'be.angry' would lack a Davidsonian event argument. Our generic predicate *máa-ń* would, therefore, have no event argument to turn to generic. My argument here is that since Yoruba treats both regular events and Kimian states (like the one in (31)) the same by using *máa-ń* (the generic predicate) to turn them from particular eventualities to generic eventualities, generic states have an ontological as well as an empirical basis. This suggests that there may be languages where Kimian states can be shown to have a hidden event argument. If my assumptions are correct, Yoruba is one of those languages. This framework also led to the ambitious suggestion that *máa-ń* is an overt pronunciation of the Gen operator.

Finally, the framework is able to handle issues relating to focus and ambiguity. Both individuals and eventualities can be put in focus. The mechanism developed here can be implemented variously to account for each case. (32a) and (33a), both repeated from (i)b and (i)d respectively in Footnote 2, demonstrate respectively how Yoruba grammatically puts generic eventualities and individuals in focus. Both sentences express the same idea; the difference is in what is put in focus. This difference is reflected in the logical forms in (32b) and (33b).

(32)   a. Gbígbó   ni   ajáá   gbó
          barking   FOC   dog.**gen**   bark
          'BARKING is what dogs do (generically/habitually)'
       b. ∃e [**barking** (e) ∧ **gen** (e) ∧ Gen x [**dog** (x) → **ag** (e, x)]]

(33)   a. Ajá   ní**í**   gbó
          dog   FOC.**gen**   bark
          'DOGS bark generically/ It is dogs that bark generically or habitually.'
       b. Gen x [**dog** (x) ∧ ∃e [**barking** (e) ∧ **gen** (e) → **ag** (e, x)]])

We can illustrate how the framework handles ambiguity with the popular example 'Typhoons arise in this part of the Pacific'. The two standard readings of this sentence are shown below:

(34)   a. Gen x [**typhoon** (x) → ∃e [**arising** (e) ∧ **gen** (e) ∧ **ag** (e, x) ∧ **in** (e, **this_part of_the_pacific**]]
       b. ∃x [**this_part of_the_pacific** (x) ∧ ∃e [**arising** (e) ∧ **gen** (e) ∧ ∃y [**typhoon** (y) ∧ **ag** (e, y)] ∧ **in** (e, x)]])

While (34a) contains both a generic individual and generic eventuality, (34b) contains a generic eventuality and concrete individual. (34a) states that the generic individual 'typhoons' is such that it generically occurs in this part of the Pacific, while (34b) states that this part of the Pacific

is such that some generic event of arising whose agent is concrete individual 'typhoons' occurs in it. Since (34b) does not contain a generic individual, Gen is not necessary. This also supports the idea that only generic individuals require the binding of the Gen operator.

7   Conclusion

The overarching claim of this paper has been that in the kind of sentences that we regard as generic, both individuals and eventualities can have generic interpretations. Perhaps the most notable proposal in this part of the system is that there is a generic individual which is distinct from kind individual and concrete (normal) individual. The distinction made between particular (concrete) and generic eventualities also appears to be conceptually and empirically supported. It was shown that Yoruba makes this clear distinction and that this distinction determines how noun phrases in the language are interpreted. Application of this framework to Yoruba generic sentences touches on two major issues related to current theories of genericity and event semantics. First, it was suggested that there is evidence in Yoruba that Kimian states do have the E-position which the generic predicate *máa-ń* applies to. Second, it was shown that the most intuitive treatment of *máa-ń* in certain constructions in the language is to give it the semantics of the Gen operator, thereby suggesting that Gen is not silent in all contexts in Yoruba after all. Adopting the framework also provides the opportunity to avoid recourse to pragmatics and an avenue to account for both subject and object arguments of generic sentences. The general prediction is that the different kinds of individuals and eventualities identified here are present cross-linguistically and that languages may have their own ways of expressing them. Further research will test some of the predictions and refine the system advanced here accordingly.

References

Ajíbóyè, O. (2005). Interpreting Yorùbá Bare Nouns as Generic. *Annual Meeting of the Berkeley Linguistics Society* 31(2).

Bach, E. (1986). The algebra of events. In P. Portner & B. H. Partee (eds.), *Formal Semantics: The essential readings*, 324-333. Oxford: Blackwell.

Baker, C.L. 1973. Definiteness and indefiniteness in English. Distributed by Indiana University Linguistics Club.

Carlson, G. (1989). On the semantic composition of English generic sentences. In G. Chierchia, B. H. Partee & R. Turner (eds.), *Properties, types and meaning*, 167- 192. Volume 2: Semantic Issues. Dordrecht: Kluwer.

Carlson, G. (1977). A unified analysis of the English bare plural. *Linguistics & Philosophy* 1:413-456.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, 6, 339‑405.

Davidson, D. (1967). The Logical form of action sentences. In N. Rescher, (ed.), *The logic of decision and action,* 81 -95. Pittsburgh: University of Pittsburgh Press.

Enç, M. (1991). The semantics of specificity. *Linguistic Inquiry 22:1-26.*

Fine, K. (1983). A defense of arbitrary objects. *Proceedings of the Aristotelian Society*, vol. 17.

Gerstner-Link, C. and M. Krifka. (1993). Genericity. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Venneman (eds.), *Handbuch der Syntax*. Berlin: de Gruyter.

Hawkins, J.A. (1978). *Definiteness and indefiniteness: A study in reference and grammaticality prediction.* New York: Humanities Press.

Higginbotham, J. (1985). On semantics. *Linguistic Inquiry* 16:547-593.

Higginbotham, J. (2000). On events in linguistic semantics. In J. Higginbotham, F. Pianesi & A. Varzi (eds.), *Speaking of Events*, 49-79. New York, Oxford: OUP.

Hilpert, M. (2014). *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.

Katz, G. (2000). Anti neo-Davidsonianism: Against a Davidsonian semantics for state sentences. In C. Tenny & J. Pustejovsky (eds*.), Events as grammatical objects*, 393-416. Stanford, CA: CSLI Publications.

Krifka M. F.J. Pelletier, G. Carlson, A. ter Meulen, G. Chierchia & G. Link. (1995). Genericity: an introduction. In G. Carlson & F.J. Pelletier (eds.), *The Generic Book*, 1–124. Chicago: The University of Chicago Press.

Link, G. (1995). Generic information and dependent generics. In G. Carlson & F.J. Pelletier (eds.), *The Generic Book*, 358–382. The University of Chicago Press: Chicago.

Maienborn, C. (2004). A pragmatic explanation of the stage-level/individual-level contrast in combination with locatives. In B. Agbayani, V. Samiian & B. Tucker (eds.). Proceedings of the Western Conference on Linguistics (WECOL) 15. Fresno, CA: CSU, 158–170.

Maienborn, C. (2005). On the limits of the Davidsonian approach: The case of copula sentences. *Theoretical Linguistics 31*, 275–316.

Maienborn, C. (2007). On Davidsonian and Kimian states. In I. Comorovski & K. von Heusinger (eds.). *Existence: Semantics and syntax*. Dordrecht: Springer, 107–130.

Maienborn, C. (2011). Event semantics. In C. Maienborn, K. von Heusinger, & P. Portner (eds.), *Semantics. An International Handbook of Natural Language Meaning*. Vol. 1, 802–829, Berlin: Mouton de Gruyter.

Mari, A., C. Beyssade and F. del Prete. (2013). Introduction. In A. Mari, C. Beyssade & F. del Prete (eds.), *Genericity*. Oxford: OUP.

Moltmann, F. (2013). On the distinction between abstract states, concrete states, and tropes. In A. Mari, C. Beyssade & F. del Prete (Eds.), *Genericity*, 293-311. Oxford: OUP.

Montague, R. (1969). On the nature of certain philosophical entities. In R. H. Thomason (Ed.), *The monist 53*, 159-194. Reprinted in Montague, R. 1974. Formal Philosophy 149-187. Yale University Press, New Haven (Conn.).

Parsons, T. (1990). *Events in the semantics of English: a study in subatomic semantics*. Cambridge, Mass.: MIT Press.

Parsons, T. (2000). Underlying states and time travel. In J. Higginbotham, F. Pianesi & A. Varzi (eds.), *Speaking of Events*, 81-93. Oxford, New York: OUP.

Pelletier, F.J. (2010). Generics: a philosophical introduction. In F.J. Pelletier (ed.), *Kinds, things, and stuff: The cognitive side of generics and mass terms*. New directions in cognitive science, 3–15. New York: OUP.

Pianesi, F. & A. Varzi. (2000). Events and event talk: An introduction. In James Higginbotham, Fabio Pianesi & Achile Varzi (Eds.), *Speaking of Events*, 3-48. New York, Oxford: OUP.

Intervention Effects in Palestinian Arabic: How question formation becomes degraded [1]

Julia Braun — Eberhard Karls Universität Tübingen

Abstract.   This paper provides novel results from semantic fieldwork on Palestinian Arabic (PA) on intervention effects. Theoretically, intervention effects can arise in simple and multiple *wh*-questions, in alternative questions and in scope marking constructions. It will be clarified why this is the case and why PA only exhibits effects in the latter two. Based on the empirical findings, it will be argued that grammaticality is not a binary phenomenon and that intervention effects, rather than turning a grammatical target sentence into an ungrammatical one, downgrade the judgements but do not necessarily make the target sentence ungrammatical. In this sense, intervention effects do exist, but the effect might not be as strong as predicted by the current theory. [2]

1   Background: diagnosing intervention effects

Semantic explanations of intervention effects (Beck 2006, 2016; Howell et al. Ms. 2017; Hohaus & Howell 2015) predict intervention effects to be *cases of ungrammaticality* caused by the interaction of different semantic operators that evaluate alternatives.

In more detail, we expect that certain constructions do not surface because a *wh*-phrase may not be separated from its associated Q-operator by an intervener[3], e.g negation, a focus sensitive operator or certain quantifiers, as exemplified in (1).

(1)   *$[Q_i$ [...[intervener [...wh-phrase$_i$...]]]]

Finding such intervention effects in a language is a special quest as it requires looking at constructions that would normally not be uttered in everyday conversations.
In order to understand intervention effects as they are observed cross-linguistically as well as their theoretical underpinning, the important questions to ask are:

(2)   What kind of ingredients do we need in order to construct intervention effects in a language?

(3)   Are there intervention effects in the language we are interested in?

(4)   Are the predictions of the semantic theory met, i.e. is there empirical evidence that intervention leads to ungrammaticality of the whole construction?

---

[1]I would like to thank Team Lambda from Tübingen and especially Anna Howell and Vera Hohaus for their amazing support. A special thanks also goes to Susan Rothstein from Bar Ilan University who helped me to finalise this paper. Thanks a lot!

[2]This paper is based on my BA-thesis on intervention effects in Palestinian Arabic and short passages from it might be used in this paper.

[3]The set of problematic interveners can be different for each language, see Beck (2006)

The plot for this paper is as follows: Section 1.1 provides a more detailed explanation of intervention effects and section 1.2 discusses ways to elicit intervention effects in Palestinian Arabic. In section 2, I will present data on PA and the results from studies on intervention effects in PA, which will be discussed in section 2.3. In the outlook in chapter 2.4, I will mention some issues that should be taken into account for future research. The appendix provides the meaning rules and lexical entries of the relevant ingredients of intervention effects.

1.1   What are intervention effects?

Intervention effects have been elicited in a large number of languages including German, Korean, Hindi, Turkish (Beck 1996), English, Japanese, French (Pesetsky 2000), Mandarin, Malayalam (Kim 2002), Dutch (Honcoop 1998), Passamaquoddy (Bruening & Lin 2001), Thai (Ruangjaroon 2002), Amharic (Eilam 2011), Samoan, Yoruba, (Howell et al. to appear). Consider the examples in (5) and (6) below:

(5)   *Minsu-man nuku-lûl po-ass-ni? (Korean)
        Minsu-only who-Acc see-Past-Q
        'Who did only Minsu see?'   (Beck, 2006, p.1)

(6)   *Wen   hat niemand wo   gesehen? (German)
        whom has nobody   where seen
        'Where did nobody see whom?'   (Beck, 2006, p.4)

Although the examples in (5) and (6) have a different syntatic structure, they are predicted to be ungrammatical because of the same underlying principle: what makes these constructions unacceptable is the way that the compositional interpretation of alternatives happens (cf. Beck 2006, 2016). Note that in both the Korean and the German question, there is a *wh*-phrase which has stayed in-situ and which is c-commanded by an intervener. As described above, this constellation separates the *wh*-phrase from its associated *Q*-operator at LF and thus leads to ungrammaticality. In a compositional account of this ungrammaticality (Beck 2006), it is assumed that *focus* like in (7) and *questions* (8) both introduce alternatives.

(7)   Only **Samira**$_F$ plays the piano. $\rightarrow$ focus on 'Samira' introduces alternatives {Beth, Ken, Ronja}

(8)   **Who** plays the piano? $\rightarrow$ question word also introduces the alternatives {Beth, Ken, Ronja}

To include these alternatives into the calculations, every node receives two different values: the ordinary semantic value and an additional alternative semantic value (Rooth 1985, 1992). Both the focused phrase as well as the question word are assumed to be evaluated by an operator. Rooth (1985,1992) assumes a $\sim$-operator to evaluate focused phrases in its scope, *wh*-phrases are evaluated by a *Q*-operator (Beck 2006). Beck's (2006) explanation of intervention effects is based on the binding properties of these two operators. Her theory shows that the $\sim$ unselectively evaluates all alternatives[4] in its scope including those that are introduced by the *wh*-phrase. Since the *Q*-operator cannot bind an undefined value relative to g,h, the whole calculation collapses and the

---

[4]Or in Beck's framework, all distinguished variables. In the following explanations, I will assume a distinguished variable framework.

structure becomes uninterpretable.[5] Applied to the example in (5), we would get the LF in (9) below.

(9) $[_{CP}\ Q_2\ [_{IP_3}\ \text{only}_C\ [_{IP_2}\ \sim C\ [_{IP_1}\ \text{Minsu}_{F_1}\ \text{saw who}\ _2\ ]]]]$

As explained in Beck (2006) and Beck & Kim (2006), since $[[\text{who}]]^g$ is undefined, $[[\text{IP}_1]]^g$ is also undefined. The $\sim$ then resets the value relative to g,h to the value relative to g which implies that $[[\text{IP}_2]]^g$ inherits the undefinedness from $[[\text{IP}_1]]^g$ and that $[[\text{IP}_2]]^{g,h}$ also gets undefined. Both $[[\text{IP}_3]]^g$ and $[[\text{IP}_3]]^{g,h}$ also inherit the undefinedness. And because $[[\text{IP}_3]]^{g,h}$ is undefined, $[[\text{CP}]]^g$ is also undefined which means that the whole structure is undefined and thus uninterpretable.

1.2   ...and how can we find them in PA?

There are four different question types that lend themselves to the elicitation of intervention effects. In the following, I will briefly mention them all and illustrate why they are good candidates to test intervention effects.

Simple wh-questions. As mentioned above, we need a *wh*-phrase that can be c-commanded by an intervener. This is obviously only possible in languages that do not front *wh*-phrases. Korean is a language that leaves its *wh*-phrases in-situ which means that intervention effects can easily be constructed. Consider again the Korean example below. The question word *nuku-lûl* is c-commanded by the intervener *man* which makes the whole question ungrammatical.[6] A corresponding LF to (10) is given in (11).

(10)   *Minsu-man nuku-lûl po-ass-ni? (Korean)
        Minsu-only who-Acc see-Past-Q
        'Who did only Minsu see?'   (Beck, 2006, p.1)

(11) $[Q_i\ ...\ [\ \sim C\ [...\text{wh}_i...]]...]$

Multiple wh-questions. In those languages that allow multiple *wh*-questions, it is also possible to elicit intervention effects. The prerequisite is that one of the *wh*-phrases must stay in-situ so that an intervener can be inserted. German is a language that allows multiple questions. However, if an intervener is inserted between the two question words, the whole structure gets ungrammatical, as illustrated in the example below.

(12)   *Wen   hat niemand wo   gesehen? (German)
        whom has nobody   where seen
        'Where did nobody see whom?'   (Beck, 2006, p.4)

(13) $[Q_i\ ...\ [\ \sim C\ [...\text{wh}_i...]]...]$

---

[5]The relevant meaning rules are provided in the appendix.

[6]For further evidence for this claim, see Beck 2006, p.3, ex.(2)

Alternative Questions. A more adventurous route to take is to use alternative questions as a means to elicit intervention effects. As observed by Beck & Kim (2006), the question in (14) is ambiguous between a polar question reading and an alternative question reading, i.e. possible answers to the former are "Yes/No" whereas the alternative question can be answered by naming one of the alternatives.

(14) Does John like Mary or Susan?

Interestingly, as soon as an intervener is inserted, the alternative question reading vanishes and only the polar question reading remains (Beck & Kim 2006), as illustated in (15). The way this phenomenon is explained is by asuming that the "intervener prevents association of the disjunctive phrase with a licensing interrogative complementizer" (Beck & Kim 2006, p.167):

(15) #Does only John like Mary or Susan? [*AltQ]   (Beck & Kim 2006, p.167)

Empirically, it is an advantage that the disjunction stays in-situ as this means that there is a distance between Q and the disjunction. A focus-sensitive item like *only* can thus be inserted as an intervener as shown in the intervention configuration below:

(16) $[Q_i...[\sim C \ [NP \ or_i \ NP]]]$

Scope Marking Structures. A fourth option is to use scope marking structures to test for intervention effects. Dayal (1994) describes scope marking structures as instances of an expletive wh-item extending the scope of a second meaningful wh-item. (17) shows an example of a scope marking structure, (18) shows an extraction structure. Both examples are taken from Dayal (1994):

(17) **Was** glaubst du, **mit wem** Maria gesprochen hat?
      what think   you with whom Maria spoken   has
      'Who do you think Maria has spoken to?'   (Dayal 1994, p.137)

(18) **Mit wem** glaubst du, dass Maria gesprochen hat?
      with whom think   you that Maria spoken   has
      'Who do you think Maria has spoken to?'   (Dayal 1994, p.137)

Scope marking constructions give us exactly what we need in order to test for intervention effects: a *wh*-item which can be c-commanded by an intervener because it has not been moved to the front.

   Interestingly, Dayal also mentions so-called "sequential scope marking" of the form in (19) and states that these sequential questions should also be regarded as scope marking constructions because "they have a *wh*-expression that seems to be semantically inert and a *wh* that can be construed as taking scope outside its syntactic domain" (Dayal 2000, p.171).

(19) What do you think? Who will Mary see?[7]   (Dayal 2000, p.171)

---

[7]There are a few reasons to believe that sequential questions are in fact not instances of scope marking: The question "What do you think?" could simply be an invitation to state your opinion. Secondly, in scope marking constructions as in (17), the order of the "subquestions" cannot be changed but the order can be changed in (19). Thirdly, it is fine to say in German: "Was glaubt er, mit wem Maria gesprochen hat?" but according to Susan Rothstein's native speaker intuitions it is weird to say "What does he think? Who will Mary see?".

To summarise the main argument of this section, consider the table in (20):

	Type of construction:	Possible candidate for testing intervention effects:
	Simple *wh*-question	✓
(20)	Multiple *wh*-question	✓
	Alternative question	✓
	Scope marking structure	✓

Theoretically speaking, intervention effects in Palestinian Arabic could be found in any of these question types. However, sometimes it is impossible to elicit the relevant data for a language due to syntactic constraints, i.e. because multiple questions do not exist in that language or because *wh*-phrases are obligatorily fronted. This is an interesting challenge and will be addressed in the next section.

2 Data

In section 2.1, I will try to convince the reader that only alternative questions and scope marking structures lend themselves to the elicitation of intervention effects in PA. I will then also introduce *focus*, which is another necessary ingredient on our way to intervention effects. Section 2.2 then provides data on intervention effects which will be discussed in section 2.3. Throughout all elicitations I used the guidelines discussed in Matthewson (2004), which means that my informants were asked to do either translation tasks or judgement tasks.

2.1 Prerequisites

According to the Encyclopedia of Arabic Language and Linguistics (A.L.L.), "Palestinian Arabic is a native language to approximately 8.5 million people" (Shahin, 2011). It is a Semitic language and belongs to the Afro-Asiatic language family (McCarus, Encyclopedia of A.L.L) Further, PA is a pro-drop language and has an SVO word order as shown below, taken from Braun (2016): [8]

---

[8]There is some debate in the literature as to what the word order in Palestinian Arabic is. I follow Shlonsky 1997 and McLoughlin 1982 in assuming an SVO word order because my participants only accepted this word order as the declarative structure. VSO order was judged to be a question by my participants.

Translation Task:
'Samira writes a book'

(21) samiira b-ti-ktib   ktaab.
      Samira IMP-FEM-write book
      'Samira writes a book.' [9]

Surface   Structure:
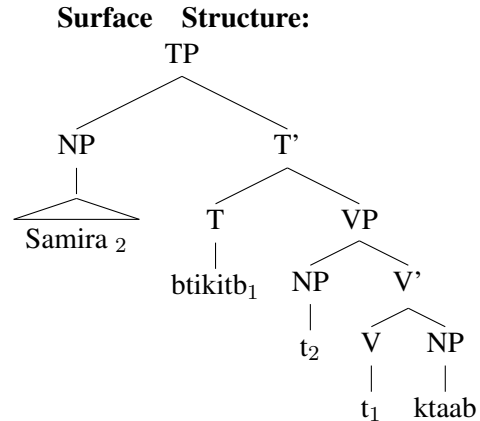
As shown in Braun (2016), when forming a polar question, the verb is moved into the head of C:

(22) b-ti-ktib   samiira ktaab?
      IMP-FEM-write Samira book
      'Does Samira write a book?'[10]

In order to form a simple *wh*-question, the question word needs to be fronted, as shown in Braun (2016) and repeated below: A surface structure of the simple *wh*-question 'What is the teacher doing?' can be seen in (24) and the corresponding translation task is given in (23).

(23) **Translation Task:** 'What is the teacher doing?'

      shu   b-t-3mal   al-mu3lm-e?
      what IMP-FEM-do the-teacher-FEM?
      'What is the teacher doing?'

(24) [CP [NP shu$_3$][C' [C bt3mal$_1$] [TP [DP al-mu3lme$_2$][T' [T t$_1$][VP [NP t$_2$][V' [V t$_1$][NP t$_3$]]]]]]]

These data show that simple *wh*-questions cannot be used to test for intervention because the question word is obligatorily fronted in PA. Another standard way to test for intervention effects is to use multiple questions. However, this option is also ruled out as illustrated in (25), taken from Braun (2016).

---

[9]For the tree structure: cf. Mohammad, 2000, p. 83; Shlonsky, 1997, p. 7f, where he assumes movement of the verb from the head of V to some functional projection and then an additional movement of the subject to the head of GP.

[10]The corresponding surface structure is:
[CP [∅] [C' [C btiktib$_1$] [TP [NP Samira $_2$][T' [T t$_1$][VP [NP t$_2$][V' [V t$_1$][NP ktaab]]]]]]]

(25)   Some of your friends (Anna, Polina and Alex) have moved to a different city and you lost
        track which of your friends now lives in which city. You've got another friend who knows
        where your friends live. You talk about Anna, Polina and Alex and then ask your friend:
        Judgement Task:

   a.   *miin bu-skun   ween?
        who IMP-live where?
        'Who lives where?'

   b.   ween bu-skun kul   waaHad?
        where IMP-live every one?
        'Where does everyone live?'

My informants uniformly rejected multiple questions which leads me to conclude that multiple
questions are in fact ungrammatical in PA and thus do not lend themselves to test for intervention
effects. Alternative questions, on the other hand, do exist. The corresponding data are presented in
(26) and taken from Braun (2016).

(26)   We went for a walk in the woods and it was very cold. We finally get back home and I ask
        you:

   a.   bitHab   qaHwe **'au** shaai?   → **PolQ**
        like(2.Ps.Sg.MASC) coffee or tea?   **possible answers are: yes / no**
        'Would you like coffee or tea?'

   b.   bitHab   qaHwe **willa** shaai?   → **AltQ**
        like(2.Ps.Sg.MASC) coffee or   tea?   **possible answers are: coffee / tea**
        'Would you like coffee or tea?'

The lexicon of PA contains two different disjunctive items, namely *willa* and *'au*. While *willa* is
reserved for alternative questions, *'au* can be used in polar questions as well as in declaratives.
This distribution of the two disjunctive items was also argued for in Winans (2013, 2015) for
Egyptian Arabic. Importantly, this means that alternative questions are valid candidates to test for
intervention effects.
    Lastly, scope marking constructions do seem to exist in PA, too. The relevant data are given in
(27) and (28).

(27)   shu raijak,   ma3 miin Hakat marijam?
        what opinion-your(MASC) with who spoke Mariam
        'Who do you think that Mariam spoke to?'

(28)   shu fikrat   monA ween raiH 3li?
        what thought Mona where went Ali
        'What did Mona think where Ali went?'

All of the data from Palestinian Arabic are summarised in the table in (29).

	Type of construction:	Possible candidate for testing intervention effects:
	Simple *wh*-question	×
(29)	Multiple *wh*-question	×
	Alternative question	✓
	Scope marking structure	✓

As explained above, there is another important ingredient of intervention effects, namely *focus*. The way that the ungrammaticality of intervention effects as in (10) or (12) is compositionally calculated is by assuming that "wh-phrases and focus make use of the same interpretational mechanism, and because of that, focus interferes with a wh-phrase in situ." (Beck & Kim 2006, p.175) There are three different focus-sensitive items in PA; namely *bas* (only), *kamaan* (also/too) and *Hataa* (even). These three items can combine with focused phrases. And example of this is given in (30), taken from Braun (2016).

(30) Salim, Ahmad and Mohammad are in a bookstore. All three of them looked at books, but in the end...
      Translation Task: 'Only Salim$_{[F]}$ bought a book.' (and no one else did so.)

      bas saliim 'ishtaraa   ktaab.
      only salim bought(3.Ps.Sg.M.) book
      'Only Salim bought a book.'

2.2 Intervention effects

Firstly, I will present an intervention effect in an alternative question, taken from Braun (2016). In a second step, I will present a small study on intervention effects in scope marking constructions.

Alternative questions. As usual, the informants were confronted with a context and asked to judge the target sentence with regard to this context. Consider the example below:

(31) Mahmud is a very nice person and he enjoys eating and drinking. He is not picky when it comes to food or drinks, so he also eats food that other people might find disgusting. Last week, you(the participant) hosted a party and you offered tea and maqlubi (an Arabic rice dish). One of those two things was very disgusting but you cannot remember which one (you did not feel well on that night, this is why you cannot remember it.) You do know, however, that Mahmud was the only guest that consumed the disgusting thing. You want to host another party next week and you want to make sure that the disgusting thing will not be offered again. You want to find out what only Mahmud ate because if you know that, then you will know what was disgusting thing.
      You ask:

Judgement Tasks:
Intervention in AltQ with *bas* as intervener:[11]

a. *bas maHmuud ’akal   maqluubi walla shirib   shaay?
  only Mahmud ate(3.Ps.Sg.MASC) maqlubi or   drank(3.Ps.Sg.MASC) tea?
  ‘Did only Mahmud eat maqlubi or drink tea?’

 No intervention in AltQ:

b. ’akal   maHmuud maqluubi walla shirib   shaay?
  ate(3.Ps.Sg.MASC) Mahmud   maqlubi or   drank(3.Ps.Sg.MASC) tea?
  ‘Did Mahmud eat maqlubi or drink tea?’

 No intervention in PolQ with *bas*:

c. bas maHmud ’akal   maqluubi?
  only Mahmud ate(3.Ps.Sg.MASC) maqlubi?
  ‘Did only Mahmud eat maqlubi?’

The corresponding LF to (31-a) is provided below:

(32)   AltQ: *[**Q** [∼ C <u>bas</u> maHmuud$_F$[’akal maqlubi **walla** shirib shaay]]]

Scope marking constructions. As mentioned before, I then designed a small study: I used minimal pairs, namely a scope marking structure without an intervener and the corresponding sentence with an intervener. These sentences were of course accompanied by different contexts that made the respective readings reasonable. An illustration of the experiment design is given in (33).

The target sentences were presented in Palestinian Arabic and the informants had to judge whether these target sentences were natural or not. In order to have an objective means to write down the judgements, I came up with a novel method which I will call the *smiley method*: as judgement tasks involve a certain "linguistic feeling" on the side of the informant, I asked the informants to colour the smiley that they connected to the feeling they had when reading the target sentence. An example of such a judgement is given in (34).

	Scope marking without intervener:	**Scope marking with intervener:**
(33)	(1a) shu fikrat Mona, ween raaH 3lii?	(1b) shu fikrat bas Mona, ween raaH 3lii?
	what thought Mona where went Ali	what thought only Mona where went Ali
	‘What did Mona think where Ali went?’	‘What did only Mona think where Ali went?’

---

[11]I would like to thank Michael Yoshitaka Erlewine for pointing out that it would be useful to elicit a simpler version of this intervention configuration where only one verb is used, i.e. "Did only Mahmud drink coffee or tea?". Unfortunately, I have not been able to elicit this yet.

Scope marking without intervener:	Scope marking with intervener:

(2a) shu bitfakir samira, miin baas jooz-ha?
    what thinks samira who kissed husband-her
    'What does S. think who her husband kissed?'
	(2b) shu bitfakir bas samira, miin baas jooz-ha?
	    what thinks only samira who kissed husband-her
	    'What does only S. think who her husband kissed?'

(3a) shu bitfakir susan, ma3 miin zaid
    what thinks Susan with who Zaid
    Tulla3 jitmashaa al-jaum?
    went.outside to-walk today
    'What does S. think with who Z. went outside
    for a walk today?'
	(3b) shu bitfakir bas Mona, ma3 miin zaid
	    what thinks only Mona with who Zaid
	    Tulla3 jitmashaa al-jaum?
	    went.outside to-walk today
	    'What does only M. think with who Z. went outside
	    for a walk today?'

(4a) shu bitfakir immha, ma3 miin Hakat
    what thinks her-mother with who spoke
    maram?
    Maram
    'What does her mother think with who Maram
    spoke?'
	(4b) shu bitfakir bas bint chaalti, ma3 miin Hakat
	    what thinks only her-cousin with who spoke
	    maram?
	    Maram
	    'What does only her cousin think with who Maram
	    spoke?'

(5a) shu bitfakir shams ay hadiye a3Taa 3mar
    what thinks Shams which present gave Omar
    liay Tifl?
    to-which child
    'What does Shams think, which present Omar gave
    to which child?'
	(no 5b) – too difficult to construct
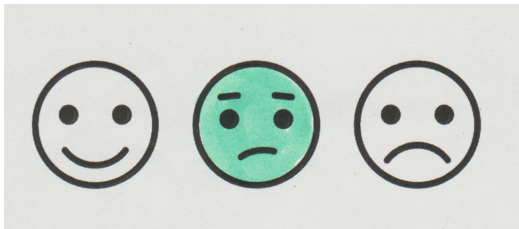
(34)

2.3 Results and discussion of the data

As indicated in (31), there was a visible difference of judgements for the alternative question with and without an intervener. While (31-b) and (31-c) were absolutely acceptable for my informants, the meaning of (31-a) seemed to be less clear. As is common in semantic field work, I only interpreted the judgements and comments of the informants as either pointing towards the grammaticality of the target sentence or towards its ungrammaticality. However, I believe that this notion of grammaticality as a binary phenomenon should be dismissed (cf. Featherston 2007, 2008), simply because a clear binary distinction of judgements is not empirically founded.

In the follow-up study on scope marking constructions, I tried to give the informants a choice between different judgements by using the *smiley method* mentioned above. I only had three informants, two of which were raised bilingually. The results provided in (35) and (36) might, however, still show a tendency:

(35)

(36)
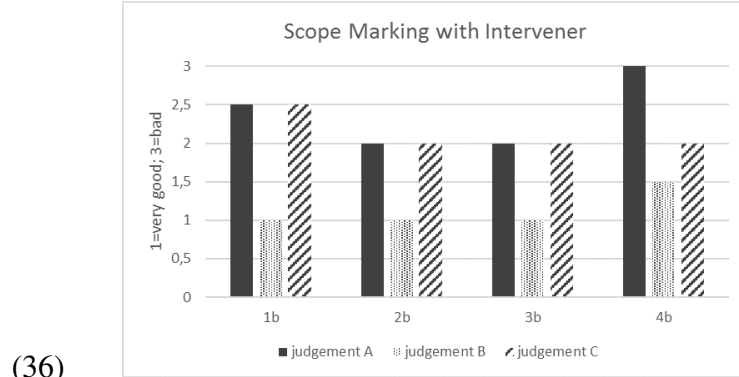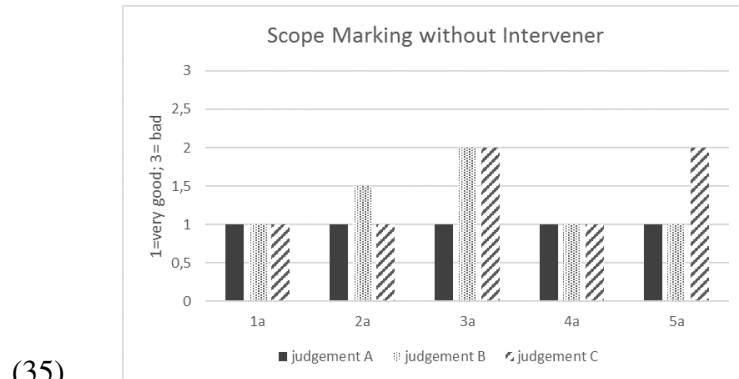
There seemed to be a rather clear effect for informant A - the monolingual Arabic speaker represented by the black bars. He judged all of the scope marking sentences without an intervener as fully acceptable whereas the corresponding sentences with an intervener were judged worse. Informant B and C were the German, Arabic bilinguals. One could assume that they were influenced by their German intuitions. However, they very clearly commented on the scope marking constructions *with* an intervener that "*bas*"/only should be left out. Comments included "to use *bas* here is really weird", "*bas* needs to be deleted", "everything is fine but we do not need the *bas*". So even though informant B and C did not give the same judgements as informant A, they did comment on the inappropriateness of *bas*. It seems, however, that the results are not as strong as expected by Beck's theory (2006). All of the scope marking constructions with an intervener should be absolutely uninterpretable. This is clearly not what I could find.

2.4   Outlook

It would be worthwhile to conduct a quantitative study with a less crude judgement scale and with more informants. Ideally, those participants should not be bilinguals in order to exclude transfer from any language other than Arabic. On top of that, in any future work on scope marking

constructions in PA, it should be checked again whether scope marking sentences do in fact exist in PA or whether they are cases of sequential questions as discussed in section 1.2.[12]

It would not be surprising to find that intervention effects are less strong than predicted by semantic explanations. Beck's (2006) theory of intervention predicts a total breakdown of compositional interpretation, however, we do more or less understand what is meant in such intervention cases, as the results in section 2.3 show. Thus, to think of intervention effects as cases that distinguish grammatical structures from ungrammatical structures (because they have an intervener) might be too strong of a claim. However, I do not want to deny that there certainly is a visible effect of intervention. I believe that grammaticality is best presented by thinking of a scale that includes different levels of acceptable and unacceptable structures as indicated in (37).

(37)
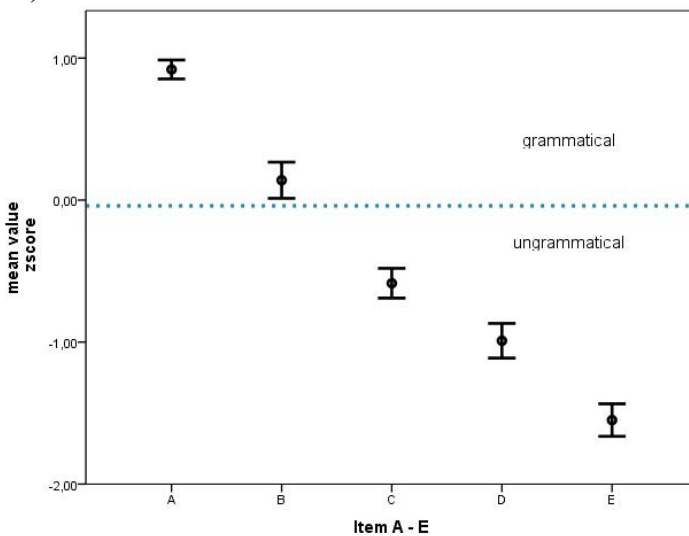
The way that I understand intervention effects is the following: the semantic calculation might derive an uninterpretable structure as discussed in Beck (2006). However, the fact that the syntax still sounds acceptable leads informants to judge the sentences as weird or downgraded but not necessarily as ungrammatical, as mention in the previous section. One consequence of this assumption would be that semantic factors of a sentence are more subtle than syntactic properties and that judgements on syntactic ungrammaticality are stronger than judgements on semantic properties. Evidence for this is the fact that my informants very clearly rejected multiple questions (which is a syntactic judgement) in PA. I believe it would be very enriching for semantic fieldwork to stop thinking about grammaticality as a binary phenomenon and to start assuming that there are different levels of grammaticality.

---

[12]Thanks to Lior Laks from the Bar Ilan University, Israel who pointed that out to me.

Appendix

(1)	(2)
Focus:	$\sim$ **operator (unselective) :**
If $\alpha = \beta_{Fi}$, then for any g,h:	If $\alpha = [\sim C\beta]$, then for any g,h:
$[\![\alpha]\!]^g = [\![\beta]\!]^g$	$[\![\alpha]\!]^g$ is only defined if
$[\![\alpha]\!]^{g,h} = $ h(i) if i is in the domain of h,	$g(C) \subseteq \{[\![\beta]\!]^{g,h}|$ h is a total distinguished
$[\![\alpha]\!]^g$ otherwise	variable assignment$\}$.
	Then, $[\![\alpha]\!]^g = [\![\beta]\!]^g$
	$[\![\alpha]\!]^{g,h} = [\![\beta]\!]^{g,\emptyset}$
(3)	(4)
question operator Q (selective):	$[\![\mathbf{only}]\!]^g =$
If $\alpha = [Q_i\beta]$, then for any g,h:	$\lambda C_{<<s,t>,t>}.\lambda p_{<s,t>}.\lambda w.\ p(w) = 1.\ \forall q\ [q$
$[\![\alpha]\!]^g = \{[\![\beta]\!]^{g,\emptyset[x/i]}|x \in D\}$	$\in C\ \&\ q \neq p \rightarrow q(w) = 0]$
$[\![\alpha]\!]^{g,h} = \{[\![\beta]\!]^{g,h[x/i]}|x \in D\}$	

References

Beck, S. (1996). *Wh-constructions and transparent Logical Form.* PhD thesis, Universität Tübingen.

Beck, S. and S.-S. Kim (2006). Intervention effects in alternative questions. *Journal of Comparative Germanic Linguistics* 9(3), 165–208.

Beck, S. (2006). Intervention effects follow from focus interpretation. *Natural Language Semantics* 14(1), 1–56.

Beck, S. (2016). Focus Sensitive Operators. In C. Féry and S. Ishihara (eds.), *Handbook of Information Structure.* Oxford: Oxford University Press.

Bruening, B. and V. Lin (2001). Discontinuous QPs and LF Interference Effects in Passamaquoddy. *Proceedings of the Semantics of Under-Represented Languages of the Americas (SULA)*, 20-28. Amherst, MA: GLSA.

Dayal, V. S. (1994). Scope marking as indirect wh-dependency. *Natural Language Semantics* 2(2), 137–170.

Dayal, V. S. (2000). Scope marking: Cross-linguistic variation in indirect dependency. In U. Lutz, G. Mller and A. von Stechow (eds.) *Wh-Scope Marking* (Linguistik Aktuell/Linguistics Today 37), 157–193. Amsterdam: John Benjamins.

Eilam, A. (2011). *Explorations in the informational component.* PhD thesis, University of Pennsylvania.

Featherston, S. (2007). Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33(3), 269–318.

Featherston, S. (2008). Thermometer judgements as linguistic evidence. In C.M. Riehl and A. Rothe (eds.), *Was ist linguistische Evidenz?*, 69–90. Aachen: Shaker Verlag.

Howell, A., V. Hohaus, P. Berezovskaya, K. Sachs, S. Erbektas, J. Braun and S. Beck (2017).

(No) Variation in the Grammar of Alternatives. Ms., Universität Tübingen.

Hohaus, V. and A. Howell (2015). Alternative semantics for focus and questions: Evidence from Samoan. In A. Camp, Y. Otsuka, C. Stabile, and N. Tanaka (eds.), *Proceedings of the 21st Meeting of the Austronesian Formal Linguistics Association (AFLA)*, 69–86.

Honcoop, M. (1998). *Dynamic Excursions on Weak Islands*. PhD thesis, University of Leiden.

Kim, S.-S. (2002). Intervention Effects are Focus Effects. *Japanese/Korean Linguistics 10*, 615–628.

Matthewson, L. (2004). On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4), 369–415.

McCarus, E.N. (2011). Modern Standard Arabic. *Encyclopedia of Arabic Language and Linguistics*, Managing Editors Online Edition: L. Edzard and R. de Jong. Consulted online 28 July 2016.

McLoughlin, L. (1982). *Colloquial Arabic (Levantine)*. London: Routledge and Kegan Paul.

Pesetsky, D. (2000). *Phrasal Movement and its Kin.* Cambridge: The MIT Press.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics* 1(1), 75–116.

Ruangjaroon, S. (2002). Thai Wh In-situ. Paper presented at the 14th Northwest Linguistics Conference, Simon Fraser University, April 6-7.

Shahin, K.N. (2011). Palestinian Arabic. *Encyclopedia of Arabic Language and Linguistics,* Managing Editors Online Edition: L. Edzard and R. de Jong. Consulted online 28 July 2016.

Winans, L. D. (2013). *Alternative Disjunctions in Egyptian Arabic*. MA thesis, University of California, Los Angeles.

Winans, L. D. (2015). Disjunction and alternatives in Egyptian Arabic. *Questions in Discourse.* Brill.

Probing the ignorance of Epistemic Indefinites: A (non)-Familiarity constraint[1]

Rahul Balusu — EFLU, Hyderabad

Abstract. Epistemic Indefinites (EI), with an added 'ignorance' component not seen with ordinary indefinites, are licensed in broadly two contexts. In one, there is ignorance of the witness of the existential claim among a plurality of referents in the domain. In the other, the witness of the existential claim can be identified, but there is ignorance about certain aspects of this witness. The first context is analysed as domain widening of some sort (Alonso-Ovalle and Menendez-Benito 2003, *et seq*). The second context is analysed as domain shifting (Aloni & Port 2010), with two identification methods at play, one required for knowledge (the ignorance component) and one used for specifying of the EI. Available identification schemes are naming, description, & ostension. We focus on the second context, and show that the Telugu EI *eed-oo/evar-oo* (which-DISJ/who-DISJ) 'some-thing/body' can be used even when the speaker has access to all 3 methods of identification, which should not be possible going by the domain shifting account. We propose a solution along the lines that if the speaker has recognized the person, *evar-oo/eed-oo* cannot be used. Similar to the familiarity theory of definiteness with discourse referents (Karttunen 1976) and 'file cards' (Heim 1983), we formalise this notion using mental referents and mental files (Recanati 2013, 2016).

1 Introduction

An Epistemic Indefinite (EI) has an added 'ignorance' component that is not present with an ordinary indefinite, as shown in (1)-(2).

(1)   a.   oka   pustakamu kinda paDindi
           one/a book   down fell
           'A book fell down.'
           Existential component: $\exists x.book(x).fell\,down(x)$
     b.   eed-oo   pustakamu kinda paDindi
           which-DISJ book   down fell
           'Some book fell down.'
           Existential component: $\exists x.book(x).fell\,down(x)$
           Modal component: The speaker does not know which book fell down.

(2)   a.   uma   naa-ku oka pustakamu iccindi
           Uma I-DAT   one book   gave
           'Uma gave me a book.'
           Existential component: $\exists x.book(x).give(Uma, x, me)$

Proceedings of TripleA 4, 79-90
Edited by Elizabeth Bogal-Allbritten and Elizabeth Coppock

b. uma naa-ku eed-oo   pustakamu iccindi
    Uma I-DAT which-DISJ book   gave
    'Uma gave me some book.'
    Existential component: $\exists x.book(x).give(Uma, x, me)$
    Modal component: The speaker does not know which book Uma gave.

The question that comes up in this context is what counts as ignorance. How much can you know and still use *eed-oo*? What licenses *some* in English? When can you absolutely not use *eed-oo*? What anti-licenses *some*? The other question is how to encode the ignorance. What is the source? Is it pragmatic, an implicature, or is it conventionalized, part of the meaning, or is it domain widening, or is it domain shifting?

EIs are common cross-linguistically. As Haspelmath (1997) notes, they involve a "(lack of) knowledge of the speaker" and he also says that this phenomenon "has received very little attention in the theoretical literature." In fact, they were first noticed by Strawson (1974) who says "that the choice of *some* rather than *a* embodies what might be called an acknowledgement or recognition of the fact that the identification supplied, though perhaps the best the speaker can do, might be regarded as inadequate to the circumstances of the case; and that the kind of identification which the choice of *some* rather than *a* indicates or suggests inability to provide (though perhaps sometimes accompanied by indifference to or unconcern about) **may be either further kind-identification or individual identification.**" (Strawson 1974: 92). He illustrates with the example given in (3).

(3)   Some V.I.P/cabinet minister/general has been shot.

Epistemic indefinites have been examined cross-linguistically in mostly Romance, Germanic and Slavic languages: English singular *some* (Becker 1999; Farkas 2002; Weir 2012), German *irgendein* (Kratzer and Shimoyama 2002; Aloni 2007; Lauer 2010; Port 2010; Aloni and Port 2013, Chierchia 2013), Spanish *algún* (Alonso-Ovalle & Menéndez-Benito 2003, 2008, 2010, 2011), Greek *kapjos* (Giannakidou and Quer 2013), French *quelque* (Jayez and Tovena 2007, 2013), Italian (un) *qualche* (Zamparelli 2007; Aloni and Port 2013; Chierchia 2013), Romanian *vreun* (Farkas 2002, 2006, Fălăuş 2009, 2011, 2012, 2014), Russian *-to* series & *-nibud* series (Geist 2008), Czech *-si* indefinites (Simík 2016), Slovak *vola-* and *si-* series (Richtarcikova 2013), Sinhala *wh-da* & *wh-hari* (Slade 2011), and the Japanese *wh- ka* indeterminates (Sudo 2010; Kaneko 2011; Alonso-Ovalle and Shimoyama 2014). They differ from each other in the environments licensed –they are sensitive to different types of knowledge. Here we will examine the Telugu EI *eed-oo/evar-oo* 'what-DISJ/who-DISJ' paired with singular NPs.

EIs vary in their properties. Some EIs don't mind reinforcement, whereas others do. Some EIs don't mind cancellation, others do. Telugu *eed-oo* doesn't like cancellation, as shown in (4).

(4)   a. oka pustakamu kinda paDindi. adi naa diary
           one/a book   down fell   that my diary
           'A book fell down. It is my diary.'
       b. eed-oo   pustakamu kinda paDindi. #adi naa diary
           which-DISJ book   down fell   that my diary
           'Some book fell down. It is my diary.'

Telugu *eed-oo* also doesn't like being reinforced, as shown in (5).

(5)   a.   oka   pustakamu kinda paDindi. adi eed-oo   naa-ku teliyadu
            one/a book   down fell   that which-DISJ I-to   know-not
            'A book fell down. I don't know which one it is.'

       b.   eed-oo   pustakamu kinda paDindi. #adi eed-oo   naa-ku teliyadu
            which-DISJ book   down fell   that which-DISJ I-to   know-not
            'Some book fell down. I don't know which one it is.'

This is in line with the generalisation in the literature that the more 'specialized' the morphologically, the more pronounced the ignorance effect.

    In this paper we first examine the proposals for explaining the ignorance component in the literature, and show how they fail with the Telugu data in §2, before attempting a formalization of our own proposal for the Telugu indefinites, in §3.

2   What is the Speaker ignorant of?

Epistemic indefinites are licensed in broadly two contexts. In one, there is ignorance of the witness of the existential claim among a plurality of referents in the domain. In the other, the witness of the existential claim can be identified, but there is ignorance about certain aspects of this witness.

2.1   Case 1: Multiple referents

The speaker doesn't know who or what the witness of the existential claim is:

(6)   a.   **Context:** Speaker is sitting with 3 others on a sofa. When he/she bends down to tie shoelaces, one of the others pulls his/her hair. The speaker doesn't see who.

       b.   (mii-loo) evar-oo   naa juTTu laageeru
            you-in   who-DISJ my hair   pulled
            'Someone (among you) pulled my hair.'

This can be captured using von Fintel's formulation for *some* (that he adapted from Dayal's *whatever*) –variation of the individuals satisfying the existential claim across the speaker's epistemic worlds.

(7)   a.   LF: some (P) (Q)

       b. $\exists w', w'' \in D_w [\{x : P(w')(x) \& Q(w')(x)\} \neq \{x : P(w'')(x) \& Q(w'')(x)\}]$ (D_w is the set of worlds compatible with the speaker's evidence in w)

This is also captured in an implicature approach (Alonso-Ovalle and Menéndez-Benito (AO-MB) 2003, 2008, 2010, 2011a; Chierchia 2006, 2013; Fălăş 2009, 2011a,b, 2014) by saying that the EI is a domain widener: The domain of existential quantification is maximal. Domain widening is costly. It signals that alternative, more specific, smaller domains are false, lead to false claims. We can illustrate with an example: Let's say the 3 others sitting on the sofa are A, B, C. Then we get the implicatures and their exhaustification as shown in (8).

(8)   a.   In all accessible worlds, Someone among A, B, C pulled the Speaker's hair.

       b.   Alternatives/Competitors:

            In all accessible worlds, Someone among A, B pulled the Speaker's hair.

            In all accessible worlds, Someone among A, C pulled the Speaker's hair.

            In all accessible worlds, Someone among C, B pulled the Speaker's hair.

            In all accessible worlds, Someone among A pulled the Speaker's hair.

            In all accessible worlds, Someone among B pulled the Speaker's hair.

            In all accessible worlds, Someone among C pulled the Speaker's hair.

But suppose there is only partial variation, and not all the alternatives are live, as shown in (9b). Here Uma does not know where Ravi is, but not all rooms are epistemic possibilities. This is not *total* variation or ignorance, only *partial* variation or ignorance.

(9)   a.   **Context:** Ravi and Uma are playing hide-and-seek. Uma knows that Ravi is hiding in the house and she knows that he is not in the bedrooms, but he could be in any of the others:

       b.   Ravi eed-oo   room-loo unnaaDu.

            Ravi which-DISJ room-in   is

            'Ravi is in some room.'

AO-MB model this partial variation scenario as an anti-singleton constraint, as shown in (10):

(10)   a. $[\![algun]\!] = \lambda f.\lambda P_{\langle e,t\rangle} : |f(P)| \succ 1 .\lambda Q_{\langle e,t\rangle} \exists x[f(P)(x) \;\&\; Q(x)]$

       b.   Domain (P) = {bedroom, bathroom, kitchen, study, dining room, living room}

       c.   f(P) = {bathroom, kitchen, study, dining, living}

       d.   In all accessible worlds, Ravi is in a room in Bath, Kitchen, Study, Dining.

       e.   Alternatives/Competitors:

            In all accessible worlds, Ravi is in a room in Bath.

            In all accessible worlds, Ravi is in a room in Kitchen. Etc.

       f.   All the alternatives are False.

2.2   Case 2: Single referent

In these cases, the speaker can identify the witness of the existential claim, by ostension, for example, as shown in (11a).

(11)   a.   **Context:** At the conference dinner, the speaker is pointing out to his/her friend a delegate who has climbed onto the table and is singing:

       b.   A: evar-oo   delegate table ekki   paaDutunnaaru

              who-DISJ delegate table climbed singing

            A: 'Some delegate is singing climbing onto the table.'

       c.   B: evaru? 'who'   A (pointing): atanu 'that person'

The use of a demonstrative, a rigid designator, picks out the *same* individual across the speaker's epistemic worlds. There is going to be no variation, the variation condition is not met, as shown in (12). But *evar-oo* is good here. So this analysis won't work.

(12) $\exists w', w'' \in D_w [\{x : P(w')(x) \& Q(w')(x)\} = \{x : P(w'')(x) \& Q(w'')(x)\}]$

Should the variation condition be scrapped? A Lewisian ontology could save the variation condition (AO&MB), as shown in (13a).

(13)   a.   Individuals exist in only one world (the real one). Cross-world identity is modelled via counterpart relations. Counterpart relations are *similarity* relations. For *x* to be a counterpart of *y*, *x* and *y* have to be substantially similar, along a certain dimension.

         b. $\exists w', w'' \in D_w [\{x : P(w')(x) \& Q(w')(x)\} \not\equiv \{x : P(w'')(x) \& Q(w'')(x)\}]$
               $\equiv$ is the counterpart relation that the context demands.

For example, if *ostension* is the counterpart relation that is available, but if the context demands *naming* as the counterpart relation, the EI can be used felicitously.

Aloni & Port (2010, 2013) base their analysis on such a contextually determined 'identification' condition, implemented in a dynamic semantics framework: "On the one hand, the indefinite is used specifically. Traditionally, this means that the speaker has someone in mind, that is, she can identify the referent of the indefinite. On the other hand, the use of an EI conveys that the speaker does not know who the referent is, that is, she cannot identify the referent of the indefinite." (Aloni & Port 2015 p.129). As they say "two identification methods are at play here: the speaker can identify on one method (for example by description) but not on another (for example naming)" and the "referents of EIs are typically identified via a method different from the one contextually required for knowledge." Finally, "the notion of a **C**ontextual **C**over-shift is the technical counterpart of this intuition." For example, the "typical situation in which EIs are used is one in which the speaker can identify the referent by description, but not by name. Another quite typical situation is one in which she can identify by name, but not by ostension." This is shown in (14).

(14)   a.   I have to meet some professor. He is the Head-of-Dept. But I don't know his name.
               Speaker-can-identify $\rightarrow$ Description, Contextually-required $\rightarrow$ Naming
         b.   I have to meet some professor. His name is Noam Chomsky. But I don't know what he looks like.
               Speaker-can-identify $\rightarrow$ Naming, Contextually-required $\rightarrow$ Ostension
         c.   Some professor is singing on the table. Over there in the corner.
               Speaker-can-identify $\rightarrow$ Ostension, Contextually-required $\rightarrow$ Naming
               German: Good; Italian: Bad

The contrast between German and Italian motivates their ranking, as shown in (15):

(15)   a.   Ostension $\succ$ Naming $\succ$ Description
         b.   In Romance, but not in Germanic, the identification method required for knowledge must be higher in order than the identification method required for specific uses of EIs

So if a referent is identified by ostension, then EIs should be infelicitous in Romance, but not in Germanic.

2.3   The problem of the Telugu EI

The properties we know about EIs are that the EI is relativized to an epistemic modal base which always obeys variation (and uniformity on one dimension): either the speaker doesn't know who the witness is among many others –a plurality of potential referents; or the speaker knows the witness by one method of identification, but not by some other(s).   An analysis to cover both properties would say that variation is caused by indeterminacy of the existential claim: variation of individuals yields non-constant reference; and variation of properties that lead to individuation which are less fine-grained than getting all the way down to the individual yield a single referent but varying on some dimension. The identification schemes that are in play are Naming, Description, and Ostension.

But these Identification Schemes fail for the Telugu EI. Consider the example shown in (16), where the Speaker has access to all three methods of identification, and yet use the EI.

(16)   a.   **Context:** Speaker can read the name-tag (with name and affiliation) of the person standing on the table and singing:

b.   Ravi Sen, HoD, EFLU ani zevar-oo professor table ekki   paaDutunnaaDu
Ravi Sen, HoD, EFLU said who-DISJ professor table climbed singing
'Some Professor called Ravi Sen, HOD, EFLU, is singing on the table.'

How can the EI be used if Naming, Description, Ostension are all available to the Speaker?

On the other hand, there are context where the Speaker cannot use the EI *eed-oo/evar-oo*, even if there is an Identification Scheme that the speaker does not have access to, as shown in (17).

(17)   a.   **Context:** Speaker sees his next-door neighbor (of whom he knows nothing, except that he lives next door) slip and fall on the road:

b.   #evar-oo   jaari   paDDaaru
who-DISJ slipped fell
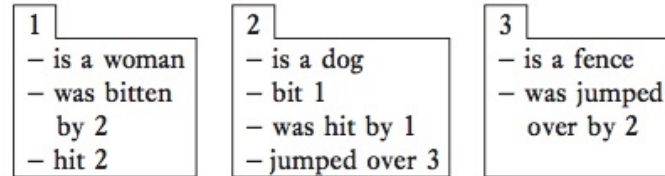'Somebody slipped and fell.'

In the above context the Speaker is more ignorant of the witness to the existential claim than in the previous context where the Speaker knew the name and description. So there is more variation, more possible worlds compatible with the speaker's epistemic state in the above example than in the previous example. Yet the example is bad. No variation story can save this situation. If the Speaker has *recognized* the person, he cannot use the EI. But what does 'recognized' mean? How do we formalize it?

3   Mental referents

Our proposal is that the Identification Scheme at work here is *familiarity*. We formalize it in the following way. We already know Discourse Referents, going back to Karttunen, and Heim (1983) and the familiarity theory of definiteness: A definite is used to refer to something that is already familiar at the current stage of the conversation. An indefinite is used to introduce a new referent. Karttunen reformulates the familiarity theory using the notion of "discourse reference". A definite NP must pick an already familiar discourse referent, whereas an indefinite NP always introduces a

new discourse referent. An NP may have a discourse referent even when it has no referent. Heim (1983) identifies Karttunen's discourse referents with "file cards", as shown in (18).

(18)   A woman was bitten by a dog. (b) She hit it. (c) It jumped over a fence

1	2	3
− is a woman	− is a dog	− is a fence
− was bitten by 2	− bit 1	− was jumped over by 2
− hit 2	− was hit by 1	
	− jumped over 3	

3.1   Recognitional files, Mental Referents, and Mental Files

In our analysis we will make use of Recanati's Mental Files (2013, 2016): To have a singular thought about an object, a subject must have a mental file that refers to the object. To have a mental file, the subject must be acquainted with its referent. Acquaintance is a relation through which a subject receives information from an object –An Epistemically Rewarding (ER) relation. Mental files are typed by acquaintance relations. Each file-type M is associated with an acquaintance relation R_M such that the referent of a file-token m of type M is the unique object o to which the subject stands in the R_M relation. The referent of a mental file is the dominant source of, rather than the object that best satisfies, the (mis)information contained in the file. Thus mental files have a non-descriptive semantics and so they are the mental analogues of referring terms.

Most mental files exist only as long as the relation they exploit contextually holds —*demonstrative files*: " when a demonstrative mode of presentation comes out of existence because the demonstrative relation on which it is based no longer holds, another relation comes to hold, in virtue of which I remember the object.": *memory demonstrative file* Through our memories of the object, we can focus our attention on it even after the perceptual encounter has ended. The memory demonstrative itself is converted into a *recognitional demonstrative* file when the object is re-encountered. Multiple exposure to that object then creates and maintains in the subject a disposition to recognize that object, via a stable **recognitional** file. The relation to the object is now 'familiarity'. Our claim is that recognitional file anti-licenses *eed-oo*.

The recognitional file doesn't have to contain perceptual information (visual, auditory, etc), it can just be some 'handle' on that entity/individual, as shown in (19):

(19)   a.   **Context:** You've been telling your mother about a new friend you made at work/university, called Ravi. She doesn't know how this Ravi looks like or sounds like. Ravi calls you at home and your mother answers. If she makes the connection with the recognitional file she has opened for Ravi then,

        nii-koosam (#evar-oo) Ravi phone ceeseeDu
        you-for   who-DISJ Ravi phone made
        'Ravi has phoned for you.'

    b.   **Context:** If she doesn't make the connection to the recognitional file:

> nii-koosam Ravi ani evar-oo   phone ceeseeDu
> you-for   Ravi said who-DISJ phone made
> 'Somebody called Ravi has phoned for you.'

3.2 Recognitional files: Human vs. others

Both *eed-oo* & *some* are sensitive to whether the referent is human or not, as shown in (20).

(20)   a.   Some professor is dancing on the table.
           b. ??Some monkey is dancing on the table.

Weir (2012) takes recourse to the 'anti-singleton' constraint of Alonso-Ovalle & Menendez-Benito (2010) to explain the contrast in examples like those in (21) and (22).

(21)   a. ??Some statue is in the middle of the square. [looking at it]
           b. ??I saw some building on my way through the desert.
           c. ??There's some letter in my mailbox [looking at it]

(22)   a.   Somebody is in the middle of the square. [looking at the person]
           b.   I saw someone on a camel on my way through the desert.
           c.   There's somebody in my room. [looking at the person]

Weir (2012) states that "A speaker uses *some NP$_{thing}$* to signal that she could not, if presented with the extension of NP, reliably differentiate the witness of the existential claim from everything else in the extension of NP. A speaker says *some NP$_{sub-kind}$* to signal that she cannot restrict to a singleton the set of subkinds within the extension of NP to which the witness of the existential claim belongs."

We can here ask the question, why do humans go down to the level of individuals, and non-humans to the kind/sub-kind level? We suggest that it is because that's how we cognitively 'parse' them and 'store' them. Humans go down to the level of individuals. Non-humans stop at the sub-kind/species level, as illustrated in the examples in (refsoup.

(23)   a.   naa soup-loo eed-oo   purugu undi
               my soup-in what-DISJ insect is
               'There's some insect in my soup.'
           b. ??naa soup-loo eed-oo   ciima undi
               my soup-in what-DISJ ant   is
               Intended: 'There's some ant in my soup.'
           c.   naa soup-loo (oka) ciima undi
               my soup-in (one) ant   is
               'There's an ant in my soup.'
           d.   swimming pool-loo evar-oo   unnaaru
               swimming pool-in who-DISJ is
               'There's someone in the swimming pool.'

The individual vs. sub-kind difference that *eed-oo* is sensitive to tracks the individual vs. sub-kind

difference in recognitional files for humans vs. non-humans. Some non-humans like books, dogs, pets, can go down to the individual level. On the other hand some humans can stop at the sub-kind level like policemen, watchmen, soldiers, etc (people in uniform).

3.2.1 Kind files

We can capture this difference by using the distinction of *natural-kind concepts*: A variety of recognitional files whose content is not an individual object. We can then posit that humans have individual recognitional files, whereas non-humans have natural-kind recognitional files

(24) a. (i) I saw a camel on the way here.
          (ii) #I saw some camel on the way here.
      b. (i) I saw a man on a unicycle on the way here.
          (ii) I saw some man on a unicycle on the way here.
      c. (i) There is some plant growing out of the wall of my room (Weir 2012).
          (ii) I saw some appliance/gadget on the kitchen table this morning.

This is similar to the *type* vs. *token* distinction in Japanese (Alonso-Ovalle & Shimoyama 2014), where *dore-ka* is for tokens and *nani-ka* is for types.

3.3 Recognitional Files: Celebrities

But in certain contexts, like with heads-of-state and celebrities, *eed-oo* is anti-licensed even where there does not seem to be a recognitional file, as shown in (25a).

(25) a. **Context:** You are watching a news channel. They are telecasting live from the UN. Suddenly there are gun-shots and commotion. The ticker flashes that the President of Peru has been shot. You say to your friend:
          Peru President-ni   shoot ceeseeru
          Peru President-ACC shoot did
          'Peru's President is shot.'

      b. **Context:** You are watching a news channel. The news is about a shooting in a school.
          America school-loo John Smith anee evar-oo   student-ni   shoot ceeseeru
          America school-in   John Smith said who-*disj* student-*acc* shoot did
          'Some student called John Smith in a school in America was shot.'

So does this mean that celebrities or V.I.Ps are more easily familiar than non-celebrities? It could also be that the notion of familiarity with a celebrity is suitably calibrated to the distance that celebrities keep from ordinary folk, or some relative metric like that. Another possibility is that these are actually Kind recognitional files and not Individual recognitional files. It could also be that we expect hearers also to be familiar with them, taking hearers into the computation of the notion of familiarity.

3.4 *Some* doesn't care for recognitional files

English *some* can be used for an entity that the speaker has a recognitional file for, as shown in (26).

(26)   a.   Ravi is someone that I know from childhood.
       b.   This curry is something I've eaten since childhood.

This is of course not possible with *eed-oo/evar-oo*, as shown in (27).

(27)   #Ravi naa-ku evar-oo   baagaa   telisinavaaDu
       Ravi me-to who-DISJ very-well known-guy
       Intended: 'Ravi is someone that I know very well.'

So *eed-oo/evar-oo* and *some* split ways here. While recognitional files anti-license *eed-oo/evar-oo*, they don't anti-license *some*.

4   Conclusion

In this paper we saw clear evidence for an Indentification Scheme beyond Naming, Description, Ostension to capture the distribution of the Telugu EI *eed-oo* —Familiarity. We made an initial attempt at grounding this identification scheme in the notion of mental recognitional files of various kinds, with humans being sensitive to Individual recognitional files and non-humans being sensitive to Kind recognitional files.

References

Aloni, Maria. 2007. Free choice, modals, and imperatives. *Natural Language Semantics*, **15**(1), 65–94.

Aloni, Maria, & Franke, Michael. 2013. On the free choice potential of epistemic and deontic modals. *From grammar to meaning: the spontaneous logicality of language. Cambridge: Cambridge University Press. Google Scholar*.

Aloni, Maria, & Port, Angelika. 2010. Epistemic indefinites and methods of identification. *Ms. University of Amsterdam*.

Aloni, Maria, & Port, Angelika. 2012. On epistemic indefinites: a note on emphatic free choice uses. *Pages 1–14 of: Proceedings of Sinn und Bedeutung*, vol. 16.

Alonso-Ovalle, Luis, & Menéndez-Benito, Paula. 2003. Some epistemic indefinites. *Pages 1–12 of:* Kadowaki, M., & Kawahara, S. (eds), *Proceedings of NELS 33*, vol. 33.

Alonso-Ovalle, Luis, & Menéndez-Benito, Paula. 2008. Minimal domain widening. *Pages 36–44 of: Proceedings of the 27th west coast conference on formal linguistics*. Cascadilla Proceedings Project Somerville, MA.

Alonso-Ovalle, Luis, & Menéndez-Benito, Paula. 2010a. Domain restrictions, modal implicatures and plurality: Spanish algunos. *Journal of Semantics*, **28**(2), 211–240.

Alonso-Ovalle, Luis, & Menéndez-Benito, Paula. 2010b. Modal indefinites. *Natural Language Semantics*, **18**(1), 1–31.

Alonso-Ovalle, Luis, & Menéndez-Benito, Paula. 2013a. Epistemic Indefinites: Are We Ignorant About Ignorance? *Programme Committee*.

Alonso-Ovalle, Luis, & Menéndez-Benito, Paula. 2013b. A Note on the Derivation of the Epistemic Effect of Spanish Algún as an Implicature. *Pages 36–49 of: Alternatives in Semantics*. Springer.

Alonso-Ovalle, Luis, & Menéndez-Benito, Paula. 2013c. Two views on epistemic indefinites. *Language and Linguistics Compass*, **7**(2), 105–122.

Alonso-Ovalle, Luis, & Shimoyama, Junko. 2014. Expressing ignorance in the nominal domain: Japanese wh-ka. *Pages 11–20 of: Proceedings of West Coast Conference on Formal Linguistics*, vol. 31.

Becker, Misha. 1999. The some indefinites. *Syntax at Sunset*, **2**, 1–13.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic inquiry*, **37**(4), 535–590.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention*. OUP Oxford.

Falaus, Anamaria. 2009. Polarity items and dependent indefinites in Romanian. *Université de Nantes dissertation*.

Falaus, Anamaria. 2010a. Alternatives as sources of semantic dependency. *Pages 406–427 of: Semantics and Linguistic Theory*, vol. 20.

Falaus, Anamaria. 2010b. New challenges in the area of semantically dependent indefinites: the Romanian epistemic constraint. *Romance Linguistics*, 287–302.

Falaus, Anamaria. 2014. (Partially) Free choice of alternatives. *Linguistics and Philosophy*, **37**(2), 121.

Farkas, Donka F. 2002. Varieties of indefinites. *Pages 59–83 of: Semantics and linguistic theory*, vol. 12.

Geist, Ljudmila. 2008. Specificity as referential anchoring: evidence from Russian. *Pages 151–164 of: Proceedings of Sinn und Bedeutung*, vol. 12.

Giannakidou, Anastasia, & Quer, Josep. 2013. Exhaustive and non-exhaustive variation with free choice and referential vagueness: Evidence from Greek, Catalan, and Spanish. *Lingua*, **126**, 120–149.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Clarendon Press Oxford.

Heim, Irene. 1983. File change semantics and the familiarity theory of definiteness.

Jayez, Jacques, & Tovena, Lucia M. 2002. Determiners and (un) certainty. *Pages 164–183 of: Semantics and Linguistic Theory*, vol. 12.

Kaneko, Makoto. 2011. DP external epistemic 'determiners' in Japanese. *Empirical issues in syntax and semantics*, **8**, 239–66.

Karttunen, Lauri. 1969. Discourse referents. *Pages 1–38 of: Proceedings of the 1969 conference on computational linguistics*. Association for Computational Linguistics.

Kratzer, Angelika, & Shimoyama, Junko. 2017. Indeterminate pronouns: The view from Japanese. *Pages 123–143 of: Contrastiveness in Information Structure, Alternatives and Scalar Implica-*

tures. Springer.

Lauer, Sven. 2010. Some news on irgendein and algún. *Pages 10–12 of: Talk presented at Workshop on Epistemic Indefinites, University of Göttingen.*

Recanati, François. 2012. *Mental files*. Oxford University Press.

Recanati, François. 2013. *Reference through Mental Files: Indexicals and Definite Descriptions*.

Recanati, François. 2016. *Mental files in flux*. Oxford University Press.

Richtarcikova, Veronika. 2013. Ignorance and indifference: Epistemic indefinites in Slovak.

Šimík, Radek. 2013. Epistemic indefinites under epistemic modals in Czech. *In: Prezentované na Potsdam Syntax-Semantics Colloquium*.

Slade, Benjamin Martin. 2011. *Formal and philological inquiries into the nature of interrogatives, indefinites, disjunction, and focus in Sinhala and other languages*. University of Illinois at Urbana-Champaign.

Strawson, Peter Frederick. 2017. *Subject and predicate in logic and grammar*. Routledge.

Sudo, Yasutada. 2010. Wh-ka pronouns in Japanese and the semantics of indeterminate pronouns. *Pages 10–12 of: Talk presented at the Workshop on Epistemic Indefinites, University of Göttingen and Lichtenberg-Kolleg*.

Weir, Andrew. 2012. Some, speaker knowledge, and subkinds. *ESSLLI 2012 Student Session*, 180.

Zamparelli, Roberto. 2008. On singular existential quantifiers in Italian. *Pages 293–328 of: Existence: Semantics and syntax*. Springer.

Egophoric attitudes and questions
in Kathmandu Newar[1]

Stephen Wechsler — The University of Texas
David Hargreaves — Western Oregon University

Abstract.   Kathmandu Newar (Sino-Tibetan) has an *egophoric* verb marking system: an *egophoric* (or *conjunct*) verb form co-occurs with first person in declaratives and second person in interrogatives. Egophoric marking is restricted to predicates of intentional action and also interacts with evidential markers. This paper examines the distribution of egophoric marking in reports of speech and attitudes, extending to this domain the analysis of egophoric marking as indicating self-ascription by the epistemic authority for the utterance. This distribution reveals that egophoric marking of a clause further introduces an implication that the epistemic authority believes the proposition denoted by the clause.

1   Introduction

In *egophoric* verb marking systems, a special form of the verb called the *egophoric* (or *conjunct*) form is found in first person statements and second person questions, while the *non-egophoric*, (or *disjunct*) form appears elsewhere. Such a system is found in the Tibeto-Burman language Newar (Nepal Bhasa) spoken primarily in the Kathmandu Valley. The characteristic interrogative flip pattern of Newar is illustrated with the past tense forms of the verb meaning 'go' shown in Table 1. The egophoric form *wan-ā* bears a special suffix and also neutralizes the perfective/imperfective distinction found on the non-egophoric forms.

	declarative	interrogative
1st person	*wan-ā*	*wã: / wan-a*
2nd person	*wã: / wan-a*	*wan-ā*
3rd person	*wã: / wan-a*	*wã: / wan-a*

Table 1: Finite past forms of Newar 'to go'. The non-egophoric forms distinguish perfective from imperfective. The macron indicates vowel quality: *ā* = low central vowel; *a* = low back vowel.

The egophoric verb form is glossed EGO in the following examples (from Hargreaves 2005, ex. (51) to (56)):

(1)   a.   jĩ: a:pwa twan-ā.
              1.ERG much drink-PST.EGO
              'I drank a lot.'

b. jĩ: a:pwa twan-a   lā.
    1.ERG much   drink-PFV.NON.EGOQ
    'Did I drink a lot?'

(2) a. chã   a:pwa twan-a.
    2.ERG much   drink-PFV.NON.EGO
    'You drank a lot.'

b. chã   a:pwa twan-ā   lā.
    2.ERG much   drink-PST.EGOQ
    'Did you drink a lot?'

(Glosses below are simplified to EGO and NON.EGO, omitting any indication of the tense/aspect. All such examples are in the past tense.). EGO marking also interacts with evidentiality, as we shall see below. In previous work we have sought to specify the semantics of the EGO marker (Hargreaves, 2005, in press; Wechsler, in press; Coppock & Wechsler, to appear). This paper continues that project, focusing now on evidence from the use of EGO in subordinate clauses expressing the contents of speech and attitude reports.

Following that earlier work, we posit that the EGO morpheme signals that the proposition denoted by its clause is the content of an attitude one has about oneself, called a *de se* attitude or a *self-ascription* (Lewis, 1979a). The role of the participant in the utterance context who self-ascribes that content is called the *epistemic authority*: roughly the speaker uttering a declarative and the addressee hearing a question, modulo the systematic effects of evidential markers. When the EGO morpheme marks a subordinate clause in Newar, we find that the self-ascriber of that clause's content must be the *reported* epistemic authority, and not the root authority. For example, in reports of attitudes, coreference with the *reported* attitude-holder (*Rām* in (3)) determines EGO-marking in the subordinate clause:

(3) a. Rām-ā: (wã)   gāka   jyā   yān-ā   dhakā: cāl-a.
    Ram-ERG 3.ERG enough work do-EGO COMP be.aware-PFV.
    'Ram$_i$ realized (lit. became aware) that he$_{i/*j}$ had done enough work.'

b. Rām-ā: (wã)   gāka   jyā   yāt-a   dhakā: cāl-a.
    Ram-ERG 3.ERG enough work do-NON.EGO COMP be.aware-PFV.
    'Ram$_i$ realized (lit. became aware) that he$_{*i/j}$ had done enough work.'

With EGO on the subordinate verb, its subject is coreferential with the matrix subject; with NON.EGO it is not. Interestingly, if the attitude verb in a sentence like (3a) is negated ('Ram didn't realize...'), the EGO marking in the subordinate clause disappears, even if the subjects are coreferential (see Section 3).

Next we review the distribution of Newar EGO marking in main clauses (Section 2) and embedded clauses (Section 3), followed in Section 4 by a summary of formal analysis in Coppock & Wechsler (to appear). Then we modify the formal analysis to account for new data from clausal complements (Section 5).

2 Semantic properties of Newar EGO-marking

As shown above, first person declarative clauses are typically EGO marked. However, when certain evidential words appear, first person subjects can cooccur with the NON.EGO form of the verb.

(4) jĩ: a:pwa twan-a   khanisā.
     1.ERG much drink-NON.EGO EVID
     'It appears I drank a lot.'

Sentence (4) would be appropriate if the speaker doesn't remember drinking, perhaps because heavy drinking wiped away any memory of the event. Without EGO, the evidential source for the information expressed in the utterance is understood to be indirect or inferential. In contrast, the EGO form is used, as in (1a), by a speaker who remembers carrying out the action described. Conversely, evidential marking can also enable an EGO-marked verb to cooccur with a third person subject, if the individual denoted by the subject is understood as the source of the report:

(5) a. syām-ã   a:pwa twan-ā   hã
         Syam-ERG much drink-EGO EVID
         'According to Syam$_i$, he$_i$ drank a lot.'

     b. wã   a:pwa twan-a   hã
         3.ERG much drink-NON.EGO EVID
         'It is said that he drank a lot.'

In sentence (5a) with EGO marking, *Syam* is understood as the source of the report. But in (5b) without EGO, the source is understood to be hearsay originating from someone other than a participant in the event.

Summarizing so far, the subject of an EGO-marked verb is the *(epistemic) authority* for the utterance (Hargreaves, in press; Hale, 1980).[2] In declarative sentences the authority is typically the speaker, in which case the subject is in first person. But authority can be deferred in evidentials, leading to third person uses. In interrogatives, the authority is the addressee of the utterance, so the subject is in second person.

EGO appears only in descriptions of intentional actions. With a first person subject, the verb *thwān* in EGO form indicates intentional kicking, but in NON.EGO form it indicates accidental bumping:

(6) a. jĩ: wa-yāta thwān-ā.
         1.ERG 3-DAT   kick-EGO
         'I kicked him/her [intentionally].'

---

[2]Other terms for notions equivalent or closely related to epistemic authority include: *commitment holder* (Krifka, 2014), *informant* (Bickel, 2008), *epistemic source* (Hargreaves 2005), *seat of knowledge* (Speas & Tenny, 2003), *locutionary actor* (Hale, 1980), *locutor* (Aikhenvald, 2004), and *judge* (McCready, 2007).

b. jĩ: wa-yāta thwān-a.
    1.ERG 3-DAT   kick-NON.EGO
    'I bumped against him/her [by accident].'

Inherently non-intentional predicates like *thyan-* 'arrive' and *thul-* 'understand' never take the EGO form. Still others, like the verb *twan-* 'drink' in example (1) and (2) above, indicate intentional action by default, hence take EGO form, but this default can be overridden with modification by *macāeka* 'unwittingly', in which case the NON.EGO form is used.

As noted in the introduction, the EGO morpheme signals the content of a *de se* attitude held by the epistemic authority. In other words, the epistemic authority knowingly *self-ascribes* that content. This property of EGO marking is best shown with subordinate clause uses. Consider the following scenario: *Syam is looking at a photo from a wild party in which someone is wearing a lampshade on his head. Syam points at the besotted partier and says to you, 'That guy drank too much'; unbeknownst to Syam, it is himself in the picture.* This scenario can be reported with the English sentence (7a), but not with the controlled infinitive in (7b):

(7) a.   Syam$_i$ said that he$_i$ drank too much. (need not be *de se*)
     b.   #Syam$_i$ claimed PRO$_i$ to have drunk too much. (*de se* only)

The Newar sentence (8) is syntactically like the English (7a), but semantically like (7b):

(8) #Syām-ã   wã   a:pwa twan-ā   dhakā: dhāl-a.
     Syam-ERG 3.ERG much drink-EGO COMP say-PERF
     'Syam$_i$ said that he$_i$ drank too much.'

Like the English control construction, (8) cannot describe the scenario above: the EGO marking forces the interpretation in which Syam knowingly self-referred.

We take this *de se*-ness to be a fundamental semantic property of EGO morphology. From this fundamental property, together with certain general assumptions about conversational pragmatics, we will derive the restriction to first person in root declaratives (unless deferred using evidentials) and second person in root interrogatives.

3 Egophoricity in embedded clauses

In the clausal complements of speech and attitude verbs, EGO-marking is determined by the *reported* authority rather than the root authority for the utterance. In reports of statements, first of all, coreference with the reported *speaker* determines EGO-marking in the embedded clause:

(9) a.   syām-ã   wã   a:pwa twan-ā   dhakā: dhāl-a.
         Syam-ERG 3.ERG much drink-EGO COMP say-PERF
         'Syam$_i$ said that he$_i$ drank too much.'
     b.   syām-ã   wã   a:pwa twan-a   dhakā: dhāl-a.
         Syam-ERG 3.ERG much drink-NON.EGO COMP say-PERF
         'Syam$_i$ said that he$_j$ drank too much.'

The reported speaker is the epistemic authority for the embedded declarative clause.

In reports of questions, coreference with the reported *addressee* determines EGO-marking in the embedded clause:

(10) Rām-ã: wa-yāta [wã: gāka jyā yān-ā / yāt-a dhakā] nyan-a.
Ram-ERG 3SG.DAT 3SG.ERG enough work do-EGO / do-NON.EGO COMP ask-PFV
'Ram asked him$_i$ whether he$_{i/j}$ (EGO/NON.EGO) did enough work.'

The reported addressee is the epistemic authority for the indirect question.

Finally, in reports of *de se* attitudes, coreference with the reported attitude-holder determines EGO-marking in the subordinate clause. This is shown in example (3) above. The reported attitude-holder is the epistemic authority for the embedded clause.

Summarizing, the subject of an EGO-marked verb in a complement clause must refer to the *reported* epistemic authority. Assuming, as we will later, that the authority is a parameter of the context, then the behavior of EGO-marking in embedded clauses indicates a kind of indexical shift, where the authority index has shifted to the embedded context. A similar authority shift has been observed for many evidential systems (Korotkova, 2016). To get a proper understanding of embedded egophoricity in Newar, we should consider further that *negated* attitude verbs reject EGO-complements:

(11) Rām-ã: (wã) gāka jyā yān-ā dhakā: cāl-a.
Ram-ERG 3.ERG enough work do-EGO COMP be.aware-PFV
'Ram$_i$ became aware (realized) that he$_{i/*j}$ had done enough work.'

(12) *Rām-ã: (wã) gāka jyā yān-ā dhakā: ma-cā:.
Ram-ERG 3.ERG enough work do-EGO COMP NEG-aware.IPV
'Ram$_i$ wasn't aware that he$_{i/j}$ had done enough work.'

In (11) EGO-marking on a verb in the complement clause indicates that the subject of the EGO-marked verb refers to the reported bearer of the attitude (*Rām*), as usual. But in (12), where the attitude verb has been negated, we find that the embedded verb cannot take the EGO form even if its subject is coreferential with the attitude bearer.

The badness of (12) can be explained as follows. The verb *cā-* 'be aware' is factive, so the content of the complement clause is presupposed. We posit that when a clause is EGO-marked, its content includes the information that the authority believed it at reference time. But that directly conflicts with (12), so the sentence is bad.

We will say that EGO-clauses must be *authorized*: the authority in the context should believe the proposition denoted by the clause. We explore some further consequences of that claim below.

4 Formal analysis

4.1 Overview

To capture the semantics of an attitude *de se*, Coppock & Wechsler (to appear) posited a perspectival *agent* parameter as a refinement on semantic content. The content of a statement, for example, is not a set of worlds but rather a set of world-agent pairs, or *centered worlds* (Quine, 1969; Lewis, 1979a). EGO-marking on a verb identifies its subject with that agent parameter. When a sentence is uttered, the authority in the utterance context, whose commitment to the centered world proposition is at issue, effectively centers those worlds on herself, thus self-identifying as the perspectival agent. As explained below, the semantic content interacts with the pragmatic theory to predict the person restrictions on the subjects of declaratives and interrogatives, as well as the exceptions due to the deferring of the authority by means of evidential words.

In this paper we address EGO-marking in complement clauses. Like Coppock & Wechsler (to appear) we assume that the Kaplanian context has a parameter for the *authority*. For EGO-marking in complements of speech/attitude verbs, the authority parameter shifts to the *reported* authority. We further posit that EGO-marking projects a belief by the authority in the centered proposition denoted by the clause. The latter assumption effectively explicates the notion of authority.

4.2 EGO marks self-ascribed content

Coppock & Wechsler (to appear) use a logical representation language that they call *Egophoric Logic* (EL). The extension of an expression of EL is relative to a model M, an assignment g, a Kaplanian context of utterance c, and an intensional index i. The Kaplanian context c is a tuple of the standard parameters for the speaker sp_c, addressee ad_c, time of utterance t_c, and location of utterance l_c— plus an authority parameter au_c:

$$c = \langle sp_c, ad_c, t_c, l_c, au_c \rangle$$

The intensional index i for a sentence contains parameters for worlds w_i and agents a_i:

$$i = \langle w_i, a_i \rangle$$

(We will add au_i and t_i later.) So the extension depends on an agent a, which serves as the perspectival center. This idea derives from the idea of centered worlds (Quine, 1969; Lewis, 1979a), and has been implemented similarly for the analysis of obligatory control by Anand & Nevins (2004) and Pearson (2012), among others, and for the analysis of evidentials by Korotkova (2016).

We define the *extension* of an EL expression ϕ relative to model M, context c, variable assignment g, and intensional index i, and write it as follows: $[\![\phi]\!]^{M,g,c,i}$. The *centered intension* of ϕ is a function from agent-world pairs to extensions, shown in (13a); the *uncentered intension* is a function from worlds to extensions, shown in (13b).

(13)   a. $[\![\phi]\!]^{M,g,c}_{\cancel{c}} = f : f(\langle a, w \rangle) = [\![\phi]\!]^{M,g,c,\langle a,w \rangle}$
         b. $[\![\phi]\!]^{M,g,c,a}_{\$} = f : f(w) = [\![\phi]\!]^{M,g,c,\langle a,w \rangle}$

So the centered intension of a sentence (wrt. a given M, g and c) will be a centered worlds proposition, and the ordinary/uncentered intension of a sentence (wrt. a given M, g, c and a) will be an ordinary possible worlds proposition.

With these tools, we can define an egophor as a form that picks out the agent at perspectival center. We designate the constant SELF as an egophor in EL. The extension of this expression with respect to agent a is a, shown in (14a). By way of contrast, the constant for a first person indexical is shown in (14b):

(14)　　a.　$[\![\text{SELF}]\!]^{M,g,c,i} = a_i$　　　　　　　　　　　　　　　　　　　(egophor)
　　　　b.　$[\![\text{I}]\!]^{M,g,c,i} = sp_c$　　　　　　　　　　　　　　　(1st person indexical)

Using these constants, the first person pronoun *jĩ:* translates into EL as I:

(15)　　**EL translation for Newari first person singular pronoun**
　　　　jĩ: ⇝ I

The EGO marker is a partial identity function on predicates that takes a predicate P and returns a predicate that holds of x if P holds of x and is defined if x is SELF, the perspectival center.

(16)　　**EL translation for Newari conjunct marker**
　　　　$\bar{a} \rightsquigarrow \lambda P_{et} . \lambda x . P(x) \wedge \partial(x = \text{SELF})$

Here ∂ can be read 'partial'; it yields undefinedness when the formula in its scope is not true (Beaver, 2001; Beaver & Krahmer, 2001). Assuming an appropriate translation for 'drank a lot', and appropriate composition rules (Function Application, etc.), the translation for *jĩ: a:pwa twan-ā* 'I drank-EGO a lot' will then be as follows:

(17)　　*jĩ: a:pwa twan-ā*
　　　　$\rightsquigarrow \text{DRANK-ALOT}(\text{I}) \wedge \partial(\text{SELF} = \text{I})$

The centered intension of the formula in (17) with respect to context c is a function that picks out the set of centered worlds $\langle a, w \rangle$ such that $a = sp_c$ and a drank a lot in w. This proposition is something that can serve as the object of an attitude *de se*, following Lewis (1979a).

4.3　Updating the discourse context with centered worlds propositions

We treat speech acts, including assertions and questions, as updates on discourse contexts (Lewis, 1979b; Ginzburg, 1996; Roberts, 1996; Farkas & Bruce, 2010, i.a.). In an assertion, the centered intension of the clause corresponding to the at-issue content will be what the *authority of the context* becomes committed to. If participant x is committed to a set of centered worlds P, then for all centered worlds $\langle a, w \rangle$ in P, x publicly commits to the belief that he or she may be a in w.

But it cannot be this *de se* belief that the addressee acquires in communication: if Syam expresses his *de se* attitude by saying to Mary, 'I drank a lot', then Mary does not come to believe that she (Mary) drank a lot. So centered worlds propositions do not directly update the the common ground. Instead the centered proposition is uncentered with the authority, and the resulting ordinary (uncentered) proposition enters the common ground.

We implement this idea with a discourse model that tracks the individual commitments of the discourse participants, namely a variant of Farkas & Bruce (2010), adapted slightly for use with centered worlds propositions. Each participant is associated with a set of *Discourse commitments*, which are sets of centered worlds. We also have a *Table*, following Farkas and Bruce, which is a stack of questions under discussion (QUD's). We assume that these questions under discussion are sets of centered worlds propositions. The *Common Ground* on the other hand is a set of non-centered, ordinary possible worlds. We also adopt from Farkas and Bruce a notion of a *Projected Set*, which is a set of projected future common grounds corresponding to different ways of resolving the issue on the table.

Speech acts are operations that update such contexts. If ϕ is *asserted* in context c, then the centered intension of ϕ is added to au_c's discourse commitments. The singleton set containing the centered intension of ϕ, $\{[\![\phi]\!]^{M,g,c}_{\phi\!\!\!/}\}$, is placed on the Table. And finally, a common ground is projected in the Projected Set which integrates the authority-uncentered intension of ϕ, $[\![\phi]\!]^{M,g,c,au_c}_{\$}$. This is an ordinary possible worlds proposition that is obtained by saturating the perspectival center parameter with the authority of the context. So when the proposal is accepted, the other discourse participants need not acquire a *de se* attitude to the centered worlds proposition.

Let us consider some examples of assertions. In (18), repeated from (1) above, EGO-marking indicates that the proposition expressed is centered around the subject, and since the subject is a first person pronoun, it is centered on the speaker. So all pairs $\langle a, w \rangle$ in the centered intension are such that $a = sp_c$. The speaker is the authority, who commits to this centered intension, and no problems will arise when we put the authority-uncentered version in the common ground.

(18) jĩː: aːpwa twan-āː.
      1.ERG much drink-PAST.EGO
      'I drank-EGO a lot'

In (19) the centered intension is centered around the addressee:

(19) *chã aːpwa twan-āː.
      2.ERG much drink-PAST.EGO
      'You drank-EGO a lot.'

But the authority is still the speaker, so this is problematic. The speaker probably does not want to self-ascribe being the addressee, and furthermore, the authority-uncentered version will be a contradictory proposition, the empty set.

The authority of the context au_c is usually but not always the speaker, as Faller 2006 and Murray 2010, i.a., have emphasized in connection with evidentials, so it is not always the speaker that becomes committed to the centered intension of ϕ. In the case of reportative evidential marking, we propose that the source of the information being reported be considered the authority of the context au_c. This individual, rather than the speaker, is committed to the content of the clause. Since evidential marking is used to indicate that the authority is someone other than the speaker, we predict that there should not be egophoric marking on the verb in the presence of evidential marking with a first person subject.

5 EGO-marked clauses must be authorized

In section 3 we saw that the subject of an EGO-marked verb in a complement clause is the *reported* epistemic authority: in reports of declaratives, the reported speaker; in reports of interrogatives, the reported addressee; and in reports of attitudes, the reported attitude holder. And we saw in (12) that negated attitude verbs reject EGO-marked complements. To account for this fact, we stipulate that EGO-clauses must be *authorized*: the authority in the context should believe the proposition denoted by the clause.

To formalize the authorization requirement, we first define a *doxastic accessibility* relation $\mathcal{R}_{dox}$: $\langle a, w \rangle$ stands in $\mathcal{R}_{dox}$ to $\langle a', w' \rangle$ iff it is compatible with what a believes in w for a to be a' in w'. We will say that an agent a *believes* $\mathcal{P}$ in world w according to M iff for $\langle a', w' \rangle$ such that $\langle a, w \rangle \, \mathcal{R}_{dox} \langle a', w' \rangle$: $\mathcal{P}(\langle a', w' \rangle)$ = T.

The *believes* relation is used to define a new logical constant AUTHORIZED in EL:

(20) $[\![\text{AUTHORIZED}]\!]^{M,g,c,i} = f : f(\mathcal{P}) = $ T iff au_c *believes* $\mathcal{P}$ in w_i at t_c according to M.

Now let us revise our EL translation for EGO:

(21) $\text{-}\overline{a} \rightsquigarrow \lambda P_{et} . \lambda x . P(x) \wedge \partial(x = \text{SELF} \wedge \text{AUTHORIZED}(\hat{}P(x)))$

Where $\hat{}$ is a Montagovian 'hat' operator giving the centered intension: $[\![\hat{}\phi]\!]^{M,c,g,\langle w,a \rangle} = [\![\phi]\!]^{M,c,g}_{\phi}$ Next we will consider how authorization works in root clauses, before turning to embedded clauses in the next section.

In a simple declarative, EGO indicates that the speaker believes herself to be the individual denoted by the subject, and believes *de se* the proposition denoted by the clause. Example (1) is repeated here:

(22) jĩ: a:pwa twan-ā
     1.ERG much drink-EGO
     'I drank-EGO a lot'

Due to the condition of authorization imposed by the EGO marker, an utterance of (22) implies that the speaker (qua authority) believes herself to be the speaker (qua referent of the subject *jĩ:* 'I') and to have drunk a lot. If a second person pronoun replaces the subject of (22) but the verb remains in EGO form, then the sentence implies that the speaker believes herself to be the addressee. Hence such a sentence is unacceptable, as shown in (19).

When evidentials defer authority to a third party then that third person subject must authorize the clause:

(23) syām-ã   a:pwa twan-ā   hã.
     Syam-ERG much drink-EGO EVID
     'According to Syam$_i$, he$_i$ drank a lot.'

Here authority is conferred upon Syam. The use of the EGO form implies that the authority (Syam) believes himself to have drunk a lot.

We saw that in root questions with second person subjects, EGO-marking is required:

(24) chã   a:pwa twan-ā   lā?
      2.ERG much drink-EGO Q
      'Did you drink-EGO a lot?'

Let us assume that the meaning of a polar question is the set containing the propositions denoted by the prejacent and its negation (Hamblin, 1958; Karttunen, 1977).

$$\llbracket ?\phi \rrbracket_{\cent}^{M,g,c} = \{ \llbracket \phi \rrbracket_{\cent}^{M,g,c}, \llbracket \neg\phi \rrbracket_{\cent}^{M,g,c} \}$$

For the question to be answerable, the addressee should believe some proposition in that set. So our new generalization is that EGO-marking on a clause denoting a set π of propositions indicates that the authority believes some proposition in π.

(25) Revised definition of AUTHORIZED:
      $\llbracket \text{AUTHORIZED} \rrbracket^{M,g,c,i} = f : f(\pi) = \text{T}$ iff there is a $\mathcal{P} \in \pi$ such that au_c *believes* $\mathcal{P}$ in w_i according to M.

The translation of EGO into EL remains the same; it is given in (21) above.

With this new definition of AUTHORIZED, the EGO marker in (24) now implies that the authority for an utterance of that sentence, namely the addressee since it is a question, believes either the prejacent or its negation. So the addressee believes either that she is the addressee (qua referent of the subject *chã* 'you') and that she drank a lot, or that she is the addressee and that she did not drink a lot. In other words, it implies that the addressee could answer the question sincerely. We cannot replace the subject with a first person pronoun and keep the EGO marking:

(26) *jĩ: a:pwa twan-ā   lā?
       1.ERG much drink-EGO Q
       'Did I drink-EGO a lot?'

Here the authorization requirement means that the addressee (qua authority) believes herself to be the speaker (qua referent of the subject pronoun). The addressee does not believe herself to be the speaker so the sentence is out.

A rhetorical question expresses an assertion, so the speaker, not the addressee, is the authority. As a result, EGO marking behaves as if the sentence were a declarative:

(27) Ji   ana wan-ā   lā?
      I.ABS there go-EGO Q
      'Did I go there? (I most certainly did not!)' (Hale, 1980, p. 100)

The EGO form means that the speaker believes herself to be the referent of the subject pronoun, and to have either gone there or not.

6 Analysis of egophoricity in embedded clauses

For clausal complements of speech and attitude predicates, the authority in the context shifts to the reported authority in the attitude/speech event described in the matrix clause. Shifting is nearly obligatory (an exception is discussed below):

(28)   a. *laksmĩ: jĩ: gākka jyā yān-ā   dhakā: siu:.
              Laksmi.ERG 1.ERG enough work do-EGO COMP know.IMPV
              'Laxmi knew that I worked enough.'
     b. laksmĩ: jĩ: gākka jyā yāt-a   dhakā: siu:.
              Laksmi.ERG 1.ERG enough work do-NON.EGO COMP know.IMPV
              'Laxmi knew that I worked enough.'

The EGO marker is not possible in (28a), despite the first person subject of that verb. While the root authority is speaker in the utterance, hence the referent of that first person subject, the authority has shifted to the reported authority, namely Laksmi.

    Following previous accounts of the shifting of person indexicals (Anand & Nevins, 2004, i.a.) and shifting of the epistemic authority for the interpretation of evidentials (Korotkova, 2016), we posit that the authority and time contextual parameters au_c and t_c are replaced with corresponding parameters of the intensional index. The utterance context is a tuple of parameters for the speaker, addressee, time of utterance, location of utterance, and authority:

$$c = \langle sp_c, ad_c, t_c, l_c, au_c \rangle$$

Our intensional index previously had world and perspectival agent parameters, to which we now add parameters for the authority and time:

$$i = \langle w_i, a_i, au_i, t_i \rangle$$

The interpretation of most sentences of Newar will not depend upon au_i or t_i. But speech and attitude reports will depend upon those parameters. The authority parameter au_i is fixed to be the reported speaker or attitude bearer, and the time t_i is the time of the reported speech event or attitude.

    A special shifting operator OP is defined for EL:

(29) $[\![\text{OP } \phi]\!]^{M,g,c,i} = [\![\phi]\!]^{M,g,c[au \mapsto au_i, t \mapsto t_i],i}$

When OP combines with a constituent ϕ, it overwrites the authority and time parameters of the utterance context with the authority and time values of the intensional index, for the interpretation of ϕ. The operator OP is a 'monster' in the sense of Kaplan (1977): it operates on the *character* of the item in its scope, the function from contexts to contents.

    This monstrous operator is encoded by the Newari complementizer *dhakā*, whose EL translation is given here:

(30)   *dhakā:* $\rightsquigarrow$ OP

The complementizer *dhakā* is a grammaticalized form of the verb *dhā-* 'to say'. As far as we know, authority-shifted complement clauses in Newar are always marked with *dhakā*.

Consider first speech reports. When *dhakā* combines with the bracketed embedded clause in (31), the authority for the interpretation of that bracketed clause is set to be the reported authority, Syam; and the time is set at the time that Syam spoke:

(31) syām-ã [wã a:pwa twan-ā] dhakā: dhāl-a.
Syam-ERG 3.ERG much drink-EGO COMP say-PERF
'Syam$_i$ said that he$_i$ drank a lot.'

The translation of (31):

(32) $\text{SAYS}(\text{SYAM}, \text{OP}([\hat{}\text{DRANK-ALOT}(x) \wedge \partial(x = \text{SELF} \wedge \text{AUTHORIZED}(\hat{}\text{DRANK-ALOT}(\text{SYAM})))]))$

The effect of AUTHORIZED is to imply that what Syam said is not just that he drank a lot, but that he believed he drank a lot.

Negating a speech report has no effect on the EGO-marking in the complement clause:

(33) Syām-ã wã a:pwa twan-ā dhakā: ma-dhā:.
Syam-ERG 3.ERG much drink-PST.EGO COMP NEG-say.IMPF
'Syam$_i$ didn't say that he$_i$ drank a lot.'

Here Syam did not say that he believed himself to have drunk a lot. There is no contradiction in (33).

However, negating an attitude verb is different:

(34) *Rām-ã: (wã) gāka jyā yān-ā dhakā: ma-cā:.
Ram-ERG 3.ERG enough work do-EGO COMP NEG-aware.IPV
'Ram$_i$ wasn't aware that he$_{i/j}$ had done enough work.'

The effect of the AUTHORIZED condition, contributed by the EGO marker, is to imply that at reference time, Ram believed himself to have done enough work. But (34) entails the opposite, so the sentence is unacceptable due to a presupposition failure.

In indirect questions that are reports of direct questions, authority shifts to the reported addressee. So EGO-marking depends on whether the subject refers to the reported addressee, namely the referent of the dative pronoun *wa-yāta* 'him':

(35) Rām-ã: wa-yāta [wã: gāka jyā yān-ā dhakā] nyan-a.
Ram-ERG 3SG.DAT 3SG.ERG enough work do-EGO COMP ask-PFV
'Ram asked him$_i$ whether he$_i$ did enough work.'

As with a direct question, let us assume that an indirect question denotes the set of propositions corresponding to possible answers (Karttunen, 1977). Given our revised definition of AUTHORIZED in (25), the EGO-marking on the verb in (35) indicates that at the time of the reported asking, the authority either believes that he did enough work, or believes that he did not do enough work.

Indirect question complements of attitude verbs are similar:

(36) laksmĩ: [su-yāta dhebā biy-ā-gu dhakā:] siu:.
 Laksmi.ERG who-DAT money give-EGO-NMLZ COMP know.IMPV
 'Laksmi$_i$ knows who she$_i$ gave the money to.'

The denotation of the constituent question 'who she gave the money to' is a set of propositions of the form 'she gave the money to *x*', where *x* ranges over people. The use of the EGO form implies that the reported authority (Laksmi) believes some proposition from that set, which is completely consistent with the meaning of (36).

However, if the attitude verb is negated, then this implication is inconsistent with the meaning, and so EGO-marking disappears from the indirect question.

(37) *Rām-ã̄: gāka jyā yān-ā dhakā (wã:) ma-siu:.
 Ram-ERG enough work do-EGO COMP 3.ERG NEG-know.IMPV
 'Ram$_i$ doesn't know whether he$_{i/j}$ did enough work.'

The use of the EGO form implies that either (i) Ram believes he did enough work or (ii) Ram believes he did not do enough work. In other words, it implies that Ram knows whether he did enough work. But this presupposition directly contradicts the sentence itself, so it is unacceptable.

We noted above that shifting to the reported authority is nearly obligatory, citing (28a). But there is at least one situation in which shifting is blocked, illustrated in (38).

(38) laksmĩ: jĩ: gākka jyā yān-ā dhakā: ma-siu:.
 Laksmi.ERG 1.ERG enough work do-EGO COMP NEG-know.IMPV
 'Laksmi didn't know that I worked enough (but I know I did).'

The sentence (38) explicitly denies that Laksmi could be the authority for the complement clause. So there is no shifting, and EGO is licensed instead by the root authority, the speaker. An utterance of this sentence implies that the speaker believed herself to have done enough work. Here the context has not shifted. Note that *siu:*, like English 'know', is a factive verb in Newar. Sentence (38) implicates that the complement clause is true, so given the maxim of quality, the speaker who utters this sentence implies that she believes she worked enough.

7 Conclusion

Following earlier work, we have analyzed EGO morphology as marking the content of a self-ascription. We have further postulated that EGO-marking also indicates that the epistemic authority believes *de se* a proposition in the denotation of the clause; when such propositions are centered, then the authority self-identifies as the agent at the center. For complements of speech/attitude verbs, the authority can (and normally does, unless blocked) shift to the authority of the reported context.

We are modeling the epistemic authority as a parameter of the utterance context, thus assimilating it to the more familiar elements of the context that are picked out by indexical expressions: speaker, addressee, time, and place of utterance. The tools developed for shifted indexicals have been repurposed for authority shifting. But the authority differs from those other parameters in an important respect. The authority is not a special target for reference; there are no expressions of

Newar that directly refer to the authority. Instead the authority is a component of the discourse pragmatic system itself: the authority is the person whose commitment to a proposition is either being made (as in an assertion) or being questioned. So it should not surprise us that shifting, while relatively rare and generally optional for the more familiar indexicals, is nearly obligatory for the authority. After all, the point of embedding a sentence under a speech or attitude predicate is to shift responsibility for it to the reported agent.

References

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

Anand, Pranav & Andrew Nevins. 2004. Shifty operators in changing contexts. In R. B. Young (ed.), *Proceedings from Semantics and Linguistic Theory XIV*, Ithaca, NY: CLC Publications.

Beaver, David. 2001. *Presupposition and assertion in dynamic semantics*. Stanford: CSLI Publications.

Beaver, David & Emiel Krahmer. 2001. A partial account of presupposition projection. *Journal of Logic, Language and Information* 10. 147–182.

Bickel, Balthasar. 2008. Verb agreement and epistemic marking: a typological journey from the Himalayas to the Caucasus. In Brigitte Huber, Marianne Volkart & Paul Widmer (eds.), *Chomolangma, Demawend, und Kasbek: Festschrift für Roland Bielmeier zu seinem 65. Geburtstag*, 1–14. Andiast, Switzerland: International Institute for Tibetan and Buddhist Studies.

Coppock, Elizabeth & Stephen Wechsler. to appear. The proper treatment of egophoricity in Kathmandu Newari. In Minyao Huang & Kasia M. Jaszczolt (eds.), *Expressing the self: Cultural diversity and cognitive universals*, Oxford: Oxford University Press.

Faller, Martina. 2006. Evidentiality above and below speech acts. http:// semanticsarchive.net/Archive/GZiZjBhO/.

Farkas, Donka & Kim Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27. 81–118.

Ginzburg, Jonathan. 1996. Interrogatives: Questions, facts, and dialogue. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 385–422. Oxford: Blackwell.

Hale, Austin. 1980. Person markers: Finite conjunct and disjunct verb forms in Newari. *Papers in South-East Asian Linguistics* 7. 95–106.

Hamblin, Charles L. 1958. Questions. *Australasian Journal of Philosophy* 36. 159–168.

Hargreaves, David. 2005. Agency and intentional action in Kathmandu Newar. *Himalayan Linguistics* 5. 1–48.

Hargreaves, David. in press. "Am I blue?": privileged access constraints in Kathmandu Newar. In Simeon Floyd, Elisabeth Norcliffe & Lila San Roque (eds.), *Egophoricity*, Amsterdam: John Benjamins.

Kaplan, David. 1977. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 267–298. Oxford: Oxford University Press.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1. 3–44.

Korotkova, Natalia. 2016. *Heterogeneity and uniformity in the evidential domain*. University of California Los Angeles dissertation.

Krifka, Manfred. 2014. Focus and contrastive topics in question and answer acts. Ms., Humboldt University of Berlin.

Lewis, David. 1979a. Attitudes de dicto and de se. *Philosophical Review* 88. 513–43.

Lewis, David. 1979b. Scorekeeping in a language game. *Journal of Philosophical Logic* 8. 339–359.

McCready, Eric. 2007. Context shifting in questions and elsewhere. In E. Puig-Waldmüller (ed.), *Proceedings of Sinn und Bedeutung 11*, 433–447. Barcelona: Universitat Pompeu Fabra.

Murray, Sarah. 2010. *Evidentiality and the structure of speech acts*. Rutgers University dissertation.

Pearson, Hazel. 2012. *The sense of self: Topics in the semantics of de se expressions*. Harvard University dissertation.

Quine, W. V. 1969. *Ontolgical relativity and other essays*. New York: Columbia University Press.

Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. In Jae-Hak Yoon & Andreas Kathol (eds.), *OSU Working Papers in Linguistics 49: Papers in semantics*, Columbus: The Ohio State University.

Speas, Peggy & Carol Tenny. 2003. Configurational properties of point of view roles. In Anna Maria DiSciullo (ed.), *Asymmetry in grammar*, 315–344. Amsterdam: John Benjamins.

Wechsler, Stephen. in press. Self-ascription in conjunct-disjunct systems. In Simeon Floyd, Elisabeth Norcliffe & Lila San Roque (eds.), *Egophoricity*, Amsterdam: John Benjamins.

How do Degrees Enter the Grammar?
Language Change in Samoan from [-DSP] to [+DSP][1]

Vera Hohaus — Eberhard Karls Universität Tübingen

Abstract.   The paper presents the result of a diachronic corpus study on Samoan, tracing a recent change in the setting of the Degree Semantics Parameter (Beck et al. 2009). We suggest that an earlier stage, Samoan had a negative setting of said parameter. Appropriation of another scalar concept then paved the way for the introduction of degrees into the grammar. Lexical and syntactic re-analysis of the directional particle *atu* ('forth, away') result in a new parameter setting.

1   Introduction

Languages vary in the semantics of gradable predicates. This variation is captured in the Degree Semantics Parameter in (1). In this paper, we argue that Samoan, a Polynesian language from the Austronesian family has recently undergone a change from [-DSP] to [+DSP].

(1)   DEGREE SEMANTICS PARAMETER [+/-DSP]:
        A language {does/ does not} have gradable predicates (type $\langle d, \langle e, t \rangle \rangle$ and related),
        i.e. lexical items that introduce degree arguments.
        (Beck et al. 2009, p. 19, no. (62))

The paper is structured as follows: The next section briefly reviews the semantics of comparison under the two possible settings of the Degree Semantics parameter. Section 3 provides an overview over the compositional semantics of comparison constructions in present-day Samoan and reviews the evidence for [+DSP]. Section 4 shows that we have evidence for positing an early stage of the language at which it was [-DSP]. We then go on to model this change in section 5. Section 6 concludes with some general remarks about the nature of language change.

2   The Degree Semantics Parameter

In this section, we will briefly review the diagnostics for [+DSP] and sketch a semantics for the comparative under both parameter settings. For more extensive discussion, see Beck et al. (2009), Bochnak (2013), Bochnak & Bogal-Allbritten (2015), and Bowler (2016). For an introduction to degree semantics, see Beck (2011).

Functional degree morphology used in a variety of comparison constructions like English (2) to (4) is indicative of a degree semantics. A differential comparative construction (DiffComp), however, requires a degree-based analysis. Let us briefly explain why.

(2) a. *C.J. is taller$_c$.*   (Contextual Comparative, ContextComp)
     b. *C.J. is four inches taller than Josh.*   (Differential Comparative, DiffComp)
     c. *C.J. is taller than 6ft.*   (Comparison with a Degree, DegComp)

(3) a. *Sam is as tall as Leo.*   (Equative, Eq)
     b. *C.J. is the tallest.*   (Superlative, Sup)

(4) a. *How tall is Josh?*   (Degree Question, DegQ)
     b. *Charlie is 5'8" tall.*   (Measure Construction, MeasC)
     c. *Leo is this tall, too.*   (Pronominal Measure Construction, PMeasC)

Under degree-based accounts, under which the lexical entry of a gradable adjective like English *tall* is (5) (or related, see, e.g. Svenonius & Kennedy 2006), the analysis of the DiffComp is straightforward, as addition and subtraction of degrees are easily definable (see also Stechow 2008). A sample lexical entry for a phrasal differential comparative operator inspired by Bhatt & Takahashi (2011) is in (6). In (7), we sketch the interpretation of (2-b) using this operator.

(5) $[\![\, tall_{[+\text{DSP}]} \,]\!] = \lambda d.\, \lambda x.\, \text{HEIGHT}(x) \geq d$ (type $\langle d, \langle e, t \rangle \rangle$)

(6) $[\![\, \text{-}er_{(\text{phrasal, differential})} \,]\!] = $
     $\lambda d.\, \lambda y.\, \lambda R_{\langle d, \langle e, t \rangle \rangle}.\, \lambda x.\, \text{MAX}(\lambda d'.\, R(d')(x) = 1) \geq \text{MAX}(\lambda d''.\, R(d'')(y) = 1) + d$

(7) $[\![\, [C.J.\ [[[\text{-}er\ 4\ \text{in}]\ [than\ Josh]]\ tall_{\langle d, \langle e, t \rangle \rangle}]] \,]\!] = 1$ iff $[\![\, \text{-}er \,]\!](4\,\text{in})(J)([\![\, tall \,]\!])(C.J.) = 1$
     iff $\text{MAX}(\lambda d.\, \text{HEIGHT}(C.J.) \geq d) \geq \text{MAX}(\lambda d'.\, \text{HEIGHT}(J) \geq d') + 4\,\text{in}$

     'The maximal degree d such that C.J. is d-tall exceeds
     the maximal degree d' such that Josh is d'-tall by at least four inches.'

Differential measure phrases are however well known to be problematic for vague predicate-approaches in the spirit of Klein (1980), which do not employ degrees (see the discusson in Stechow 1984a,b). A simplified degreeless lexical entry for a gradable adjective is in (8). With respect to a contextually provided comparison class C (a set of individuals), the adjective partitions its domain into two sets, those considered tall and those who are considered not tall. We can then define a phrasal comparative operator as in (9), which generates the interpretation in (10) for a simple comparative.

(8) $[\![\, tall_{[\text{-DSP}]} \,]\!]^C = \lambda x.\, x$ counts as tall with respect to C (type $\langle e, t \rangle$)

(9) $[\![\, \text{-}er_{(\text{phrasal})} \,]\!] = \lambda y.\, \lambda R_{\langle e, t \rangle}.\, \lambda x.\, R(x) = 1\ \&\ R(y) = 0$

(10) $[\![\, [C.J.\ [tall\ [\text{-}er\ [than\ Josh]]]] \,]\!]^C = 1$ iff $[\![\, tall \,]\!]^C(C.J.) = 1\ \&\ [\![\, tall \,]\!]^C(J) = 0$
      iff *C.J.* counts as tall with respect to C & *J* doesn't count as tall in C

Considering the semantics for the comparative in (9) and the compositional interpretation in (10), it is however unclear how to analyze the meaning contribution of a differential measure phrase. Vague-predicate approaches are thus only suitable if a language lacks DiffComp.

3 Comparison constructions in present-day Samoan

Crucially, present-day Samoan (PDS) has DiffComp and is [+DSP] (Hohaus 2010, 2012, 2015). In the terminology of Stassen (1985)'s typology, the comparative construction in PDS is a particle comparative that allows for ContextComp, DegComp, and DiffComp, as illustrated in (11) to (13).[2]

(11) *E   umi **atu** Malia.*   (ContextComp)
     TAM tall DIR. Mary
     "Mary is taller."

(12) *E   umi **atu** Malia i   le lima futu.*   (DegComp)
     TAM tall DIR. Mary PREP. the five foot
     "Mary is taller than 5ft."

(13) *E   umi **atu** Malia i   le lua inisi i   lo   lona   uso.*   (DiffComp)
     TAM tall DIR. Mary PREP. the two inch PREP. COMP. POSS.3.sg. sister
     "Mary is two inches taller than her sister."

Outside of comparatives, the directional particle *atu* also combines with motion predicates like *alu* ('to go') to indicate movement away from a contextually determined location like the starting point of the event or the position of the speaker. An example is in (14).

(14) *Ua   alu **atu** Sina.*
     TAM(inch.) go DIR. Sina
     'Sina has just left.'

Villalta (2007) and Hohaus (2010, 2012, 2015) conclude from their analyses of data like (11) to (13) that Samoan employs a degree semantics. Predicates like *umi* ('tall') are thus analyzed like their English counterparts, see (15-a). We follow Hohaus (2015) in analyzing the particle *atu* in comparatives (below referred to as comparison-*atu*, as opposed to motion-*atu* in (14)) as a contextual comparative operator with an optional differential degree argument, as (15-b).[3]

(15) a. $[\![\,umi\ (\text{'tall'})\,]\!] = \lambda d.\,\lambda x.\ \text{HEIGHT}(x) \geq d$ (type $\langle d, \langle e, t \rangle \rangle$)

      b. $[\![\,\text{comparison-}atu\,]\!]^c = \lambda R_{\langle d,\langle e,t\rangle\rangle}.\,\lambda x.\ \text{MAX}(\lambda d.\,R(d)(x) = 1) > c$
          $[\![\,\text{comparison-}atu_{(\text{differential})}\,]\!]^c = \lambda d.\,\lambda R_{\langle d,\langle e,t\rangle\rangle}.\,\lambda x.\ \text{MAX}(\lambda d.\,R(d)(x) = 1) \geq c + d$

      c. $[\![\,\text{MAX}\,]\!] = \lambda D_{\langle d,t\rangle}.\ \iota d\,[\forall d' \in D : d \geq d']$

Under this analysis, the compositional interpretation of the contextual comparative in (11) proceeds along the lines of (16), where the context provides some degree c as the comparison standard.

(16) $[\![\,[\text{Mary } [\text{tall}_{\langle d,\langle e,t\rangle}\ \text{comparison-}atu]]\,]\!]^c = 1$ iff

---

[2]Abbreviations used in glosses include COMP. = marker of comparison standard, CON. = connective, DEM. = demonstrative, DIR. = directional particle, EMPH. = emphatic marker, FOC. = focus marker, inch. = inchoative, neg. = sentential negation, pl. = plural, HUM. = prefix for counting humans, POSS. = possessive, PREP. = preposition, PRN. = pronoun, sg. = singular, and TAM = tense-aspect marker.

[3]Under this analysis, the *i lo*-phrase in (13) is not an argument of the comparative operator but is treated as a frame setter that indirectly manipulates the contextual standard of the comparison (see Hohaus 2015, pp. 118-127).

$$[\lambda R_{\langle d,\langle e,t\rangle\rangle}.\,[\lambda x.\ \text{MAX}(\lambda d.\ R(d)(x) = 1) > c]](\lambda d.\ \lambda x.\ \text{HEIGHT}(x) \geq d)(M) = 1$$
$$\text{iff MAX}(\lambda d.\ \text{HEIGHT}(M) \geq d) > c$$

> 'The maximal degree d such that Mary is d-tall
> exceeds some contextually provided height degree.'

Even though Samoan employs a degree semantics under this analysis, it differs from other [+DSP] languages like English in two respects, (i), the inventory of comparison constructions, and (ii), the interpretation of the unmarked form of the gradable predicate.

The inventory of comparison constructions. Samoan lacks functional morphology for other degree constructions like Eq and Sup and does not allow for DegQ, MeasC and PMeasC (Beck et al. 2009; Villalta 2007; Hohaus 2010, 2012, 2015). Relevant examples are provided in (17) to (18). See the references cited for grammatical alternatives to all these constructions without the degree predicate.

(17)   a.   *E   umi   tutusa Malia ma   Ioane.   (*Eq)
                TAM same(pl.) Mary and   John
                (Intended:) 'Mary is as tall as John.'

           b.   *E   aupito   umi Miriama.   (*Sup)
                TAM absence.of+next tall   Miriam
                (Intended:) 'Miriam is the tallest.'

(18)   a.   *'O   le   ā   umi Malia?   (*DegQ)
                FOC. the what tall   Mary?
                (Intended:) 'How tall is Mary?'

           b.   *E   umi Malia le   lima futu.   (*MeasC)
                TAM tall Mary PREP. the   five   foot
                (Intended:) 'Mary is five foot tall.'

           c.   *E   umi foi   Malia lea.   (*PMeasC)
                TAM tall   also Mary   DEM.sg.
                (Intended:) 'Mary is this tall, too.'

The interpretation of bare adjectives. Hohaus (2015, pp. 118-147) additionally reports variation in the interpretation of the morphologically unmarked (positive) form of the degree predicate. Compare the judgments for English and Samoan in (19) and (20).

(19)   a.   <u>Picture context</u>: A picture displaying five very lean men and two obese men, John and Joseph. It is clear that Joseph however weighs even more than John.
           b.   #E   puta Ioane.   c.   E   puta Iosefo.   d.   {John/Joseph} is fat.
                TAM fat   John   TAM fat   Joseph
                'John is the fattest.'   'Joseph is the fattest.'

(20)   a.   <u>Picture context</u>: A photo of two basketball players, Mary (6ft) and John (6ft 2in).
           b.   #E   umi Malia.   c.   E   umi Ioane.   d.   {Mary/John} is tall.
                TAM tall   Mary   TAM tall   John
                'Mary is the tallest.'   'John is the tallest.'

Hohaus (2015) suggests to capture this variation by positing a covert operator for the positive form in Samoan that is essentially a superlative, (21). The compositional interpretation of (20) is sketched in (22). Crucially, under this analysis, Mary needs not be considered tall against some standard of tallness derived from the comparison class, but her height is required to exceed that of everyone else in the comparison class, which in this case is only John.

(21) $[\![\text{Op}]\!]^C = \lambda R_{\langle d, \langle e,t \rangle \rangle}.\, \lambda x.\, \forall y\, [y \in C\ \&\ y \neq x$
       $\rightarrow \text{MAX}(\lambda d.\, R(d)(x) = 1) > \text{MAX}(\lambda d'.\, R(d')(y) = 1)]$

(22) $[\![[\text{Mary } [[\text{tall}_{\langle d, \langle e,t \rangle \rangle}\ \text{Op}]]]\!]^C = 1$ iff
       $[\![\text{Op}]\!]^C (\lambda d.\, \lambda z.\, \text{HEIGHT}(z) \geq d)(M) = 1$ iff
       $\forall y\, [y \in C\ \&\ y \neq M \rightarrow \text{MAX}(\lambda d.\, \text{HEIGHT}(M) \geq d) > \text{MAX}(\lambda d'.\, \text{HEIGHT}(y) \geq d')]$

       'For all individuals y in the contextually provided comparison set C but Mary, the maximal degree d such that Mary is d-tall exceeds the maximal degree d' such that y is d'-tall.'

Further support for this analysis comes from the behavior of additives like *fo'i* ('also'). In the context provided in (23), both Mary and Temukisa are quite tall. It is also true that Temukisa is taller than Mary. Crucially, though, in PDS, I cannot follow up on such a comparison by saying that Mary is tall, unlike in English.

(23)   a.   <u>Context</u>: Mary's height: 185 centimeter, Temukisa's height: 190 centimeter
      b.   *E   umi **atu** Temukisa i   lō   Malia. #'Ae e   umi fo'i Malia.*
          TAM tall DIR. Temukisa PREP. COMP. Mary   but TAM tall also Malia
          'Temukisa is taller than Mary. But Mary is also tall.'

Interim summary.   To summarize this brief overview of the grammar of comparison in PDS: The availability of a DiffComp provides evidence that Samoan is [+DSP] and has a degree semantics. In terms of degree operators, we have argued that PDS has a contextual comparative operator, which optionally takes a differential degree argument, and a covert superlative-like operator, which is used in the interpretation of the positive form of the degree predicate. These key ingredients of the grammar of comparison in PDS are summarized again below.

(24)   The grammar of comparison in present-day Samoan:
      a.   parameter setting: [+DSP]
          $[\![\text{(gradable predicate)}]\!] \in D_{\langle d, \langle e,t \rangle \rangle}$
      b.   inventory of degree operators in functional lexicon:
          $[\![\text{comparison-}atu]\!]^c = \lambda R_{\langle d, \langle e,t \rangle \rangle}.\, \text{MAX}(\lambda d'.\, R(d')(x)) > c$
          $[\![\text{comparison-}atu_{\text{(differential)}}]\!]^c = \lambda d.\, \lambda R_{\langle d, \langle e,t \rangle \rangle}.\, \lambda x.\, \text{MAX}(\lambda d.\, R(d)(x) = 1) \geq c + d$
          $[\![\text{Op}]\!]^C = \lambda R_{\langle d, \langle e,t \rangle \rangle}.\, \lambda x.\, \forall y\, [y \in C\ \&\ y \neq x$
          $\rightarrow \text{MAX}(\lambda d.\, R(d)(x) = 1) > \text{MAX}(\lambda d'.\, R(d')(y) = 1)]$

Both, the fact that PDS lacks any other comparison constructions apart from the particle comparative and the fact that the particle *atu* may occur with degree as well as motion predicates, raise the question whether this comparative construction and thus the [+DSP] setting are a recent innovation.

4 Comparison constructions in Early Written Samoan

Data indicative of such a change in parameter setting come from three sources, early descriptions of the language (section 4.2), synchronic variation in acceptability (section 4.3), and a quantitative study on a corpus of written texts from the 19th century and on a corpus from the 21st century (section 4.4). The next section provides some historical background on the language.

4.1 Historical background

The Samoan archipelago was settled approximately 3,500 years ago.[4] The written record of the islands however starts much later, in 1834, when missionaries from the London Missionary Society develop a writing system for the language based on the Latin script. The first bible translation, a dictionary and a brief grammatical description all were published in 1862 by George Pratt, followed by a first wave of other publications on and in the Samoan language. We will focus here on the period until the early 20th century, when the islands – as a result of the conflicting colonialist interests of the United Kingdom, Germany and the United States (and much civil unrest) – were partitioned into two protectorates, the Germany colony of Western Samoa and an American protectorate. We will refer to the language in this period as Early Written Samoan (EWS).

4.2 Early grammars

Crucially, none of the philological descriptions of EWS mentions the particle comparative described above (see also Stassen 1985, pp. 330-331). Neffgen (1903, pp. 7-8) even explicitly comments on the unavailability of English-like comparison constructions: „Eine eigentliche Komparation kennt der Samoaner nicht… Mitunter gebraucht der Samoaner überhaupt keine Steigerung, besonders dann, wenn keine Verwirrung oder Verwechslung entstehen kann."[5]

EWS appears to have used a conjoined comparative construction, a hallmark of degreeless languages. Funk (1893, p. 3) characterizes this construction as follows: „Bei der Komparation wird das Eigenschaftswort nicht gesteigert, sondern es werden zur Bildung des Komparativs zwei Adjektiva, von denen das eine das Gegenteil des anderen bedeutet, im Positiv zusammengestellt."[6] Pratt (1862, p. 8) writes: "Comparison is generally affected by using two adjectives, both in a positive state not in itself, but in comparison with the other." He provides the example in (25).

(25) *E   lelei lenei, a e   leaga lela.*
      TAM good this   but TAM bad   that
      'This is good, but this is bad.' (Pratt 1862, p. 8)

We conclude from this survey of the philological literature that EWS lacked the particle comparative construction of PDS that provides a crucial diagnostic for categorizing Samoan as [+DSP].

---

[4]See *A History of American Samoa* (Honolulu: Bess Press, p. 21) by the Amerika Samoa Humanities Council.

[5]Translation: "The Samoan does not have a comparison proper. Frequently, the Samoan does not use any comparison morphology, especially when no confusion or misinterpretation can occur."

[6]Translation: "For a comparison, the adjective does not change. Instead, two adjectives of which one means the opposite of the other are put together in the Positive."

The particle comparative is first mentioned in Holmer (1966, p. 27), who reports that "…it is generally stated that the comparative degree in the Polynesian languages is expressed by placing any of the local adverbs *ake* ('up') or *atu* ('away, beyond') after the adjective." As a dispreferred strategy, conjoined comparatives are however reported as late as Marsack (1975, p. 66), who provides the example in (26). The example is also relevant because it shows that in their bare, positive form, predicates like *laitiiti* ('small') are not evaluative (in the sense of Rett (2015)), that is, they do not make reference to an independent standard for size.

(26) a. Even in the case of giant vessels like the Queen Mary and the Aquitania this construction would be used. To indicate that the 85,000-ton Queen Mary is bigger than the 45,000-ton Aquitania, a Samoan of the old school would say:

b. *'Ua   telē le Queen Mary, 'ua   la'itiiti le Aquitania.*
TAM(inch.) big the Queen Mary TAM(inch.) small   the Aquitania
'The Queen Mary is big, the Aquitania is small.'

Later grammatical descriptions do not mention the conjoined comparative anymore (Hunkin 1992, Mosel & Hovdhaugen 1992, Mosel & So'o 1997), suggestion that it fell out of use.

4.3 Synchronic variation

If the particle comparative (and the loss of the conjoined comparative construction) is an innovation that dates back no more than a century, we might expect to see this change reflected in synchronic variation in acceptability when it comes to conjoined comparatives (see, e.g., Roberts & Roussou 2003, p. 236). And indeed, the construction is not consistently accepted by speakers of PDS. We discuss two exemplary cases here, the acceptability judgments pertaining to (27) and (28).

In an informal questionnaire study with six speakers, the target sentence in (27) was judged acceptable only by three speakers. One speaker rejects the sentence, while the remaining two find it only maybe acceptable. One such participant suggests to use the particle comparative instead. The example in (28) is an item from a questionnaire study with nine speakers, of which six found the conjoined comparative acceptable in this context. Two participants judged the target sentence unacceptable and one wasn't sure as to its acceptability. The latter as well as a speaker who had judged the conjoined comparative acceptable additionally expressed a preference for the particle comparative in their comments.

(27) a. <u>Context</u>: The tallest family members I have are my aunt Sulu and my brother Alofa. Both are quite tall. Sulu's height is 6'2". Alofa is even 6'4" tall.

b. *E   maualuga Alofa, e   puupuu Sulu.*
TAM high   Alofa TAM small   Sulu
'Alofa is tall, Sulu is small.'

(28) a. <u>Picture context</u>: A Lego family with two children, Mary who is indicated to be three years of age and John who is indicated to be five years of age.

b. *E   matua Ioane ae laitiiti Malia.*
TAM old   John   but young Mary
'John is old, but Mary is young.'

We hypothesize that the observed variation in the acceptability of conjoined comparatives in PDS is a result of language change from [-DSP] to [+DSP], a change that was triggered (as we will spell out in section 5) by the introduction of the particle comparative. This change made the conjoined comparative construction superfluous.

4.4 The corpus study

Under this hypothesis, the particle comparative is a recent innovation in the language. Given the facts in PDS, it is conceivable that comparison-*atu* diachronically derived from motion-*atu*. The available philological literature reviewed in section 4.2 plausibly suggest that EWS constitutes a language stage at which the particle had not yet taken on this double duty, but was used only with motion predicates. In this section, we report the results of a corpus study in which we compare the frequency of comparison-*atu* in a corpus of EWS to its frequency in a corpus of PDS. The results strongly suggest that EWS had not yet acquired comparison-*atu*.

Methodology. For the study, we constructed a corpus of Early Written Samoan from seven texts published between 1849 and 1900. The only criteria for inclusion in the corpus were the publication date (as early as possible) and the accessibility of the text. Table 2 provides an overview over the corpus. Note that four of the texts are translations into Samoan. In all those cases, second language speakers of Samoan were in charge of the translations, but, judging from the reports of several missionaries (see, e.g., Lundie Duncan 1846; Turner 1861, 1884), native speakers were usually extensively involved in the preparation of the manuscript. In the absence of any other publications in the language from the mid-19th century, we felt that the inclusion of those texts in the corpus was nevertheless warranted.

short reference	description
Mataio (1849)	Gospel of Matthew from a translation of the New Testament, published by the British and Foreign Bible Society in London
Kenese (1862)	Book of Genesis from the first Bible translation, published by the British and Foreign Bible Society in London
Violette (1875)	translation of a popular German religious textbook by Ignaz Schuster
Pratt (1890)	translation of popular fables by missionary George Pratt
Stübel (1896)	collection of 111 short Samoan stories collected between 1889 and 1891 by German diplomat Oscar W. Stübel
Fraser (1898)	five Samoan stories and songs from the archives of missionary George Pratt originally collected by missionary Thomas Powell
Sierich (1900)	collection of Samoan stories by German lawyer F. Otto Sierich

Table 1: Corpus of Early Written Samoan

Unfortunately, comparatives are not a highly frequent construction. For English, Hohaus, Tiemann & Beck (2014) find an average of approximately five comparatives per 1,000 utterances in a corpus study of adult care giver speech. A low proportion of comparison-*atu* in the EWS corpus can thus only be a meaningful indicator of language change if it is considerably lower

than the frequency with which we find it in present-day Samoan. For purposes of comparison, we therefore also constructed a corpus of PDS built from eight short books for children and young adults, published between 2004 and 2017 (Salesa 2004, Wai 2012a,b, Carlie & Chu Ling-So'o 2013, Wai 2014, Va'afusuaga 2016). The corpus additionally included 854 posts from the online edition of the Samoana section of the *Samoan Observer*[7] and from the *Samoa Times*[8] published between December 2016 and October 2017. The posts included a large number of sermons (and were thus close in genre to some of the texts in the EWS corpus).

We conducted a manual search of the EWS corpus for occurrences of *atu*, which were then categorized for use as motion-*atu*, comparison-*atu*, or as unclear. Unclear cases comprise cases of structural ambiguity in which the context did not allow us to decide between motion-*atu* and comparison-*atu* as well as cases where the translation of the sentence was unclear. Within the PDS corpus, we performed a computerized search of the news items and a manual search of books for occurrences of *atu*, which were then again categorized according to the above classification.

Results and Statistical Analysis. Tables 2 and 3 report the results of the search on the respective corpora. While there are no clear cases of comparison-*atu* in the EWS corpus among the more than 3,500 occurrences of the particle, comparison-*atu* accounts for 7.22 percent of the 2,007 hits in the PDS corpus. An example of motion-*atu* from the EWS corpus is in (29). Examples of motion-*atu* as well as comparison-*atu* from the PDS corpus are in (30) and (31).[9] (See (34) below for an example categorized as unclear.)

source	# total	# motion	# comparison	# unclear
Mataio (1849)	723	721	0	2
Kenese (1862)	821	820	0	1
Violette (1875)	1002	994	0	8
Pratt (1890)	155	155	0	0
Stübel (1896)	954	953	0	1
Fraser (1898)	0	0	0	0
Sierich (1900)	76	76	0	0
total	3,731	3,719 (99.68%)	0 (0.00%)	12 (0.32%)

Table 2: Results of search in the EWS corpus

source	# total	# motion	# comparison	# unclear
book publications	48	40 (83.33%)	8 (16.67%)	0 (0.00%)
newspaper items	1,959	1,809 (92.39%)	136 (6.95%)	13 (0.66%)
total	2,007	1,849 (92.13%)	144 (7.22%)	13 (0.65%)

Table 3: Results of search in the PDS corpus

---

[7]The *Samoan Observers* is an Apia-based newspaper with a daily print edition. Its online edition is available at <http://www.samoaobserver.ws/>. Last accessed November 16, 2017.

[8]The *Samoa Times* is an Auckland-based weekly community newspaper whose online edition is available at <http://www.samoatimes.co.nz/>. Last accessed November 16, 2017.

[9]The latter examples are from T. Aruna Loiani, "Olaga tausi aiga o le tina ia Ana," *Samoa Observer* (July 24, 2017) and V. Maiava & S. Sanerivi, "Fa'aoga tatau lou taule'ale'a," *Samoa Observer* (January 4, 2017), respectively.

(29) *ua*   [*o* ***atu***] *Leaga ma le   fanau i   le   po...*
      TAM(inch.) go(pl.) Leaga and   the offspring PREP. the night
      'Leaga and his family had left during the night...' (Stübel 1896, p. 23)

(30) *O   le tasi o ona alo   o loʻo   [malaga **atu**] i Savaii.*
      FOC. the first of 3sg. child(pol.) TAM(ipfv.) travel   DIR. to Savaii
      'His oldest child is travelling to Savaii.'

(31) *E   leai lava   se nofoaga e   [sili **atu**] nai loʻo   Samoa,...*
      TAM neg. EMPH. any dwelling TAM good DIR. from COMP. Samoa
      'There's absolutely no better place to live than Samoa,...'

The observed difference in proportion of comparison-*atu* between EWS and PDS is highly significant ($p < .0001$, two-tailed Fisher's exact test). We interpret this result to mean that EWS lacked comparison-*atu* and thus the particle comparative.

Interim Summary. Taken together, the early descriptions of the language, the synchronic variation in the acceptability of conjoined comparatives, and the results of the corpus study allow for a characterization of EWS as in (32).

(32)   The grammar of comparison in Early Written Samoan (EWS):
   a.   parameter setting: [-DSP]
   b.   inventory of degree operators in functional lexicon: ∅

So far then, Samoan is the only language for which we have evidence for a change in the [+/- DSP] parameter setting. This change appears to have taken place within the last one hundred years.[10] In the next section, we will analyze the individual developmental steps, lexical and syntactic, that are necessary to go from the grammar of EWS in (32) to the grammar of PDS in (24). For this analysis, a certain type of examples classified as unclear in our corpus search will be relevant, which is why we discuss them below. We will then move on to section 5.

Unclear Corpus Examples. While the number of unclear cases is extremely low in both corpora, we consider a certain type of ambiguous example relevant for the compositional analysis of the development of comparison-*atu* from motion-*atu*. Several occurrences of the particle were classified as unclear because of surface ambiguities that were not resolved by the immediate context. Those ambiguities arise when a verb is followed by both, an adverb and the directional particle as in (34). Here, the surface syntax is compatible with the two underlying structures in (33).

---

[10]Given the data discussed in this paper, we however cannot provide a more precise timeline of this change.

(33) a. [[V Adv] motion-*atu*]
      b. [V [Adv comparison-*atu*]]

(34) a. <u>Context</u>: The pharaoh learns about the beauty of Abram's wife Sarai.
      b. *Ona* [*agalelei* **atu**] *lea*   *o*   *ia*   *ia*   *Aperamo*. . .
          CON. treat+well DIR. DEM.sg. FOC. PRN. PREP. Abraham
          'So he [= the pharaoh] treated Abram well. . .' (Kenese/ Gn. XII:16)

In (34), under the first structure, the pharaoh's treatment is directed towards Abram (motion-*atu*)[11], and he treats him well (or better than anybody else, depending on the interpretation of the positive form). The second structure would generate a reading where the pharaoh – as a consequence of learning how beautiful Sarai is – treats her husband better than before. Crucially, the relevant bible chapter under our reading is compatible with both interpretations. We will argue that this type of surface ambiguities was essential for the introduction of comparison-*atu* into the grammar.

5   Modeling the change in parameter setting

Our starting point for this section is the semantics of motion-*atu* as used in EWS and PDS (section 5.1). An analysis of the changes involved in the development of a degree semantics in Samoan will have to address three questions: Which lexical and syntactic changes underlie the development of comparison-*atu* from motion-*atu*? How does a predicate of type $\langle d, \langle e, t \rangle \rangle$ develop from type $\langle e, t \rangle$? How does this change bring about the covert operator that PDS employs for the interpretation of the positive form? As we have not yet looked at the interpretation of the positive form in the EWS corpus, our focus in this paper will be on the first question (section 5.2).

5.1   Directed motion in EWS and PDS

Building on Hohaus (2010, 2012), we adopt an analysis of motion-*atu* under which it operates on paths. We can conceptualize paths as sequences of locations $(l_0, l_1, l_2, ..., l_n)$ whose length exceeds 1 (but see also Cresswell 1978, Piñón 1993, Krifka 1998, a.o.). Verbs of motion like *alu* ('to go') do not only describe a manner of motion (*i.e.* walking, as opposed to running), but also relate an entity to its movement paths (type l). We will be working with the lexical entry in (35).

(35) $[\![\, alu \, (\text{'go'}) \,]\!] = \lambda p. \, \lambda x. \, x$ walks along p (type $\langle l, \langle e, t \rangle \rangle$)

Intuitively, in examples like (36), motion-*atu* introduces an additional requirement of directionality on the path described by a motion predicate: It requires of every non-initial location $l_1, ..., l_n$ in the movement sequence that it be further away from the sequence's initial location l_0 than the next lower ranked location in the sequence, l_{n-1}, see also the figure. The paths that make a sentence with motion-*atu* true thus consist of totally ordered sets of locations. We adopt the lexical entry in (37), under which motion-*atu* quantifies off the path argument of the verb.

---

[11]This type of use of motion-*atu*, where the motion path is somewhat more abstract, is very frequent in Samoan (see also Hohaus 2010). The most common usage is probably with verbs of speech like *fai* ('say') and *tali* ('reply').

(36) *Ua   alu **atu** Sina.*
TAM(inch.) go DIR. Sina
'Sina has just gone away.'
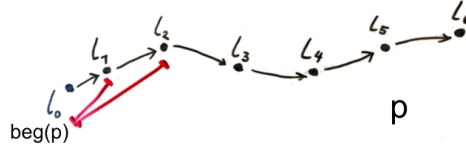
(37) $[\![\,\text{motion-}atu\,]\!] = \lambda R_{\langle l,\langle e,t\rangle\rangle}.\,\lambda x.$ (type $\langle\langle l,\langle e,t\rangle\rangle,\langle e,t\rangle\rangle$)
$\exists p\,[R(p)(x) = 1 \ \& \ \forall l\,[l \in p \ \& \ l \neq \text{beg}(p) \rightarrow l >_{\Re} \text{pre}(l)]]$

'For any $R \in D_{\langle l,\langle e,t\rangle\rangle}$ and $x \in D_e$, $[\![\,\text{motion-}atu\,]\!](R)(x) = 1$ iff there is a path p that makes $R(p)(x)$ true, and for every non-initial location l in this sequence of locations, l is ranked higher with respect to $\Re$-scale than its immediate predecessor location.'

For any location l, the underlying order $\Re$ is defined as a distance measure from the beginning of the path, DISTANCE$(l, \text{beg}(p))$. The beginning of a path $\text{beg}(p)$ is the location l such that there is no other location that precedes l. For any location l, $\text{pre}(l)$ returns the next lower ranked location in the movement sequence.

Applied to our example sentence, this analysis derives the truth conditions in (38). The sentence is true if and only if Sina walked further and further away from where she started. Somewhat more precisely, for every non-initial location in the path that Sina walked, Sina's distance to the beginning of the path increases compared to its immediate predecessor in the sequence.

(38) $[\![\,[\text{Sina }[\text{walk motion-}atu]]\,]\!] = 1$ iff
$[\![\,\text{motion-}atu\,]\!](\lambda p.\,\lambda x.\,x \text{ walked along } p)(S) = 1$ iff
$\exists p\,[S \text{ walked along } p \ \& \ \forall l\,[l \in p \ \& \ l \neq \text{beg}(p) \rightarrow l >_{\Re} \text{pre}(l)]]$

The directed movement path which motion-*atu* describes shares all important structural properties of a degree scale. We would like to suggest that this shared conceptual structure allowed for a lexical re-analysis of motion-*atu* (type $\langle\langle l,\langle e,t\rangle\rangle,\langle e,t\rangle\rangle$) as a comparative operator of type $\langle\langle d,\langle e,t\rangle\rangle,\langle e,t\rangle\rangle$. We spell this idea out in the next section.

5.2   Lexical and syntactic steps in the development of comparison-*atu*

How can we model the development of comparison-*atu* from (37), both in terms of lexical semantics and in terms of syntax? Degree comparison shares all the conceptual properties of directed movement as defined in (37). To see this, consider (39) again. The difference between Mary and the contextually provided comparison standard c on the height scale is a directed path (of degrees, rather than locations in space): The comparison standard is its beginning (and Mary's height is its end point). Given the nature of the height scale, it is indeed also true that every non-initial element of this set of degrees is further away from the beginning of this differential measurement path than the element that immediately precedes it. We would like to suggest that motion-*atu* in EWS therefore allowed for a generalization from locations in a directed movement path to degrees: Both are abstract entitites that are totally ordered. Such a transfer might have resulted in a lexical entry for comparison-*atu* along the lines of (40).

(39) *E   umi **atu** Malia.*
TAM tall DIR. Mary
"Mary is taller."
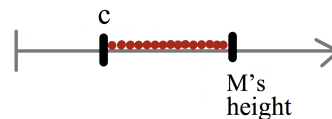
(40)   step 1a: type transfer from locations to degrees

For any $R \in D_{\langle d, \langle e,t \rangle \rangle}$ and $x \in D_e$, $[\![\,\text{comparison-}atu\,]\!]^c(R)(x) = 1$ iff
$\exists D\,[D = \{d : d \geq c\ \&\ d \leq \text{MAX}(\lambda d.\, R(d)(x) = 1)\}\ \&\ |D| > 1$
$\&\ \forall d''\,[d'' \in D\ \&\ d'' \neq c \rightarrow d >_\Re \text{pre}(d)]]$

Let's dissect this lexical entry a little: The movement path of motion-*atu* is the set of difference degrees D, our red dots in the diagram in (39). Just like paths are sequences of a length greater than 1, this set must have a cardinality greater than 1. The counterpart to the beginning of the path here is the contextually provided degree c, the standard of the comparison.

As degrees are by definition elements of a linearly ordered set, the ordering requirement expressed by the last conjunct in (40) will always be met. We can therefore re-write (40) as (41-a), which is equivalent to (41-b). We remain agnostic here as to when in this developmental sequence the differential degree argument that PDS comparison-*atu* allows for was introduced.

(41)   step 1b: simplification

a.   For any $R \in D_{\langle l, \langle e,t \rangle \rangle}$ and $x \in D_e$, $[\![\,\text{comparison-}atu\,]\!]^c(R)(x) = 1$
iff $\exists D\,[D = \{d : d \geq c\ \&\ d \leq \text{MAX}(\lambda d.\, R(d)(x) = 1)\}\ \&\ |D| > 1]$

b.   For any $R \in D_{\langle l, \langle e,t \rangle \rangle}$ and $x \in D_e$, $[\![\,\text{comparison-}atu\,]\!]^c(R)(x) = 1$
iff $\text{MAX}(\lambda d.\, R(d)(x) = 1) > c$

While the conceptual similarities between directed motion and comparison made the shift from locations to degrees in the semantics of motion-*atu* possible in the first place, the introduction of comparison-*atu* must have occurred in tandem with a change in the lexical semantics of predicates like *umi* ('tall'). It would however also have required that the language provide the right kind of syntactic environment to support both lexical changes.

Sentences like (42) from above, categorized as unclear in the EWS corpus, provide this kind of environment because of their potential surface ambiguity.

(42)   *Ona* [*agalelei* **atu**] *lea   o   ia   ia   Aperamo…*
CON. treat+well DIR. DEM.sg. FOC. PRN. PREP. Abraham
'So he [= the pharaoh] treated Abram well…' (Kenese/ Gn. XII:16)

The string consisting of a motion verb, a gradable predicate and the directional particle supports a re-analysis from motion-*atu* to comparison-*atu*, as sketched in (43).

(43)   step 2: syntactic re-analysis of surface ambiguities
[[(verb) (adverb$_{\langle e,t \rangle}$)] (motion-*atu*)] $\leadsto$ [(verb) [(adverb)$_{\langle d, \langle e,t \rangle \rangle}$ (comparison-*atu*)]]

Crucially, though, comparison-*atu* operates on degree scales and requires that its first argument be of type $\langle d, \langle e,t \rangle \rangle$. The addition of comparison-*atu* to the functional lexicon of Samoan will thus have been accompanied by a change in parameter setting from [-DSP] to [+DSP]. We don't believe that this change was "an abrupt change in grammars, reflecting a new parameter setting" (Lightfoot 1997, p. 171), but rather would have been a result of decomposition of a degreeless predicate into a predicate of type $\langle d, \langle e,t \rangle \rangle$ and the covert superlative-like operator that we find in PDS.

6   Concluding remarks

In Samoan, then, degrees enter the grammar through two well-known mechanisms of language change, borrowing from another domain and syntactic re-analysis of a potentially ambiguous surface string. Typological data suggest that borrowing from the spatial domain is very frequent in comparison constructions across languages (Stassen 1985). Interestingly, however, we are not aware of languages where the comparative is modeled after the temporal domain.

The resulting transition from [-DSP] to [+DSP] raises some interesting questions about the nature of semantic change. Language change is often argued to be "cyclical change" (Gelderen 2016, p. 4) or to resemble a spiral. The change from [-DSP] to [+DSP] however seems to be directional. Roberts & Roussou (2003) suggest to capture such directionality as markedness of one parameter setting over another. They also hypothesize that parametric change is mediated through the introduction of new functional heads. In the case of the transition from [-DSP] to [+DSP], this functional head is the degree head, comparison-*atu*.

References

Beck, S. (2011). "Comparison Constructions." In: *Semantics: An International Handbook of Natural Language Meaning*. Ed. by C. Maienborn, K. von Heusinger & P. H. Portner. Vol. 2. Berlin: De Gruyter, 1341–1389.

Beck, S., S. Krasikova, D. Fleischer, R. Gergel, S. Hofstetter, C. Savelsberg, J. Vanderelst & E. Villalta (2009). "Crosslinguistic Variation in Comparison Constructions." *Linguistic Variation Yearbook* 9: 1–66.

Bhatt, R. & S. Takahashi (2011). "Reduced and Unreduced Phrasal Comparatives." *Natural Language and Linguistic Theory* 29 (3): 581–620.

Bochnak, R. (2013). "Crosslinguistic Variation in the Semantics of Comparatives." PhD thesis. Chicago: University of Chicago.

Bochnak, R. & E. Bogal-Allbritten (2015). "Investigating Gradable Predicates, Comparison, and Degree Constructions in Underrepresented Languages." In: *Methodologies in Semantic Fieldwork*. Ed. by L. Matthewson & R. Bochnak. Oxford: Oxford University Press, 110–134.

Bowler, M. (2016). "The Status of Degrees in Warlpiri." In: *Proceedings of TripleA 2*, 1–17.

Carlie, E. & A. Chu Ling-So'o (2013). *'O le ketapila matuā fia'ai*. Wellington: Huia Publishers.

Cresswell, M. J. (1978). "Prepositions and Points of View." *Linguistics and Philosophy* 2 (1): 1–41.

Fraser, J. (1898). "Folk-Songs and Myths from Samoa." *The Journal of the Polynesian Society* 7 (1): 15–29.

Funk, B. (1893). *Kurze Anleitung zum Verständniß der Samoanischen Sprache*. Berlin: Mittler und Sohn.

Gelderen, E. van (2016). "Cyclical Change Continued: Introduction." In: *Cyclical Change, Continued*. Ed. by E. van Gelderen. Amsterdam: John Benjamins, 3–17.

Hohaus, V. (2010). "The Semantics of Motion Verbs and Comparison in Samoan." Magisterarbeit. Tübingen: Eberhard Karls Universität Tübingen.

—— (2012). "Directed Motion as Comparison: Evidence from Samoan." In: *Proceedings of Semantics of Under-Represented Languages of the Americas (SULA) 6*. Ed. by E. Bogal-Allbritten, 335–348.

—— (2015). "Context and Composition: How Presuppositions Restrict the Interpretation of Free Variables." PhD thesis. Tübingen: Eberhard Karls Universität Tübingen.

Hohaus, V., S. Tiemann & S. Beck (2014). "Acquisition of Comparison Constructions." *Language Acquisition: A Journal of Developmental Linguistics* 21 (3): 215–249.

Holmer, N. M. (1966). *Oceanic Semantics: A Study in the Framing of Concepts in the Native Languages of Australia and Oceania*. Uppsala: Lundeqvistska Bokhandeln.

Hunkin, G. A. L. (1992). *Gagana Samoa: A Samoan Language Coursebook*. Auckland: Polynesian Press.

Klein, E. (1980). "A Semantics for Positive and Comparative Adjectives." *Linguistics and Philosophy* 4 (1): 1–45.

Krifka, M. (1998). "The Origins of Telicity." In: *Events and Grammar*. Ed. by S. Rothstein. Dordrecht: Kluwer Academic Publishers, 197–235.

Lightfoot, D. (1997). "Catastrophic Change and Learning Theory." *Lingua* 100: 171–192.

Lundie Duncan, M. (1846). *Missionary Life in Samoa*. Edinburgh: William Oliphant and Sons.

Marsack, C. C. (1975). *Samoan*. 4th ed. Norwich: The English Universities Press.

Mosel, U. & E. Hovdhaugen (1992). *Samoan Reference Grammar*. Oslo: Scandinavian University Press.

Mosel, U. & A. So'o (1997). *Say it in Samoan*. Canberra: Pacific Linguistics.

Neffgen, H. (1903). *Grammatik der samoanischen Sprache nebst Lesestücken und Wörterbuch*. Wien: Hartleben.

Piñón, C. P. (1993). "Paths and Their Names." In: *Proceedings of the Annual Meeting of the Chicago Linguistics Society (CLS) 29*. Ed. by K. Beals, G. Cooke, D. Kathman, S. Kita, K.-E. McCullough & D. Testen. Vol. 2. Chicago, 287–303.

Pratt, G. (1862). *Samoan Dictionary with a Short Grammar of the Samoan Dialect*. Matautu: London Missionary Society's Press.

—— (1890). *Tala Faatusa Mai Atunuu EseEse Ua Faa-Samoaina E Palate: Fables from Many Hands Translated into the Samoan Dialect*. London: London Missionary Society.

Rett, J. (2015). *The Semantics of Evaluativity*. Oxford: Oxford University Press.

Roberts, I. G. & A. Roussou (2003). *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

Salesa, D. (2004). *Mauaina o o tātaou tu'ā (Discovering Our Ancestors)*. Wellington: Learning Media.

Schuster, I. & L. Violette (1875). *O Tala Filifilia mai Tusi Paia mai le feagaiga tuai ma le feagaiga fou*. Freiburg: B. Herder.

Sierich, F. O. (1900). *Samoanische Märchen*. Leiden: Internationales Archiv für Ethnographie.

Stassen, L. (1985). *Comparison and Universal Grammar*. Oxford: Basil Blackwell.

Stechow, A. von (1984a). "Comparing Semantic Theories of Comparison." *Journal of Semantics* 3 (1-2): 1–77.

—— (1984b). "My Reaction to Cresswell's, Hellan's, Hoeksema's and Seuren's Comments." *Journal of Semantics* 3 (1-2): 183–199.

—— (2008). "Topics in Degree Semantics: Degrees." Lecture Notes. Tübingen: Eberhard Karls Universität Tübingen.

Stübel, O. (1896). "Samoanische Texte unter Beihülfe von Eingeborenen gesammelt und übersetzt." *Veröffentlichungen aus dem königlichen Museum für Völkerkunde* 4 (2-4): 54–246.

Svenonius, P. & C. Kennedy (2006). "Northern Norwegian Degree Questions and the Syntax of Measurement." In: *Phases of Interpretation*. Ed. by M. Frascarelli. Berlin: De Gruyter, 133–162.

Turner, G. (1861). *Nineteen Years in Polynesia: Missionary Life, Travels and Researches in the Islands of the Pacific*. London: John Snow.

—— (1884). *Samoa: A Hundred Years Ago and Long Before*. R. & R. Clark: Edinburgh.

Va'afusuaga, J. (2016). *O le meaālofa mo Ana*. Auckland: little island press.

Villalta, E. (2007). "Komparative im Samoan." Handout. Tübingen: Eberhard Karls Universität Tübingen.

Wai, F. L. (2012a). *Aso Fa'ailogaina 'o le Teuila (Teuila Festival)*. Auckland: Read Pacific.

—— (2012b). *'O Lo'u Aso Muamua 'i le A'oga (My First Day at School)*. Auckland: Read Pacific.

—— (2014). *Aso Sa Pa'epa'e (White Sunday)*. Auckland: Read Pacific.