

# Selection from “Some Syntax, Semantics, and Pragmatics of Evidentiality in English” \*

Joshua Bowles  
Utah Valley University

Draft  
May 3, 2010

## Abstract

I introduce a general sketch of the formalization of some of the syntax, semantics, and pragmatics of evidentiality in English and Tuyuca. I present here a first step at integrating some of the proposals for capturing the highly constrained nature of the semantics of evidentiality, while at same time trying to explain the morphosyntactic variability of the phenomenon. I use a Minimalist type of syntax to build my model, and begin to extend this into an algebraic group theory definition of the notion *evidentiary base*; the latter is intended to be compatible with the notion of a modal base.

## 1 Introduction

The study of evidentiality is quickly becoming a popular area of study for formal semantics and pragmatics. The close relation it shares with epistemic modality and the complex interactions it has with tense and aspect make it a suitable phenomenon for elaborating detailed analysis of semantic and pragmatic issues that necessitate the formalization for contexts and situations. Evidentiality is also an interesting point of investigation for the syntax of the left periphery, the cartographic domain of functional hierarchies, and information structure in general. It is an ideal phenomenon for exploring interface conditions between semantics, syntax, and pragmatics.

I begin with some abstract considerations, mapping out general ideas and proposing structures and features that are supported by lengthy studies of evidentiality (though here I do not make direct reference to many of

---

\*This work is based on research done in the past few years on Eastern Tukanaon languages of the Northwest Amazon, specifically Tuyuca. The application of that research to English data is an outgrowth of insights gleaned from that work. I owe a debt of gratitude to Lyle Campbell for comments, direction, and advice on earlier work. I also want to thank Aniko Csirmaz for taking an interest and for many detailed discussions on previous work. Ed Rubin deserves credit for unwittingly getting me started on evidentials in the Minimalist Program, thank you. This material is new and the ideas are my own—none of those mentioned are associated with shortcomings in what follows.

them; these include works such as Aikhenvald (2004), Chung (2005), Garrett (2001) and Faller (2002)). I then introduce tentative English data to support the claim that evidentiality can be composed of specific features, in particular as an illocutionary *Force* discourse operator. Related to the hypothesis of a *Force* operator is the problem of formalizing context of discourse: for this I sketch a path for formalizing context through a formal definition of an *evidentiary base*; which is intended to be compatible with a modal base.

## 2 Cross-linguistic View in a Minimalist Setting

The general range of morphosyntactic structures responsible for  $\eta$ -interpretations<sup>1</sup> is striking. What I am assuming here is that it may be the case that  $\eta$ -interpretations are possible in ALL languages, but the particular linguistic forms or structures responsible for such interpretations are parameterized. In some languages, for example Dutch or English, these interpretations are compositionally built through particular Mode, Aspect, and clause interactions. In others, for example Estonian or Turkish, one finds aspectual forms in particular clausal constructions that serve as  $\eta$ -morphemes (i.e., they are homophonous with aspect as the result of historical reanalysis; see Campbell (1991) and Harris & Campbell (1995) for Estonian reanalysis and Johnson & Utas (2000) for Turkish). Still yet in others, most famously Tuyuca and neighboring Upper Northwest Amazonian languages, there are  $\eta$ -morphs used directly for the purposes of  $\eta$ -interpretations that no longer function in a verbal, aspectual, or mood/modal way.<sup>2</sup> In one of the most complex of these latter systems (the Tuyuca paradigm), one finds five distinct evidential morphs, each of which is fused with either present or past tense and also inflects for sentential subject agreement (i.e., only in the case where the subject of the clause is also the speaker does the inflectional agreement overlap with the speaker). Additionally, there are distinct future tense evidentials and interrogative evidentials. The following examples are taken from Bowles (2009b), which is based on data found in Barnes (1984, 1994, 1996) and Karn (1976).

- (1) Pakí yái sǎ̃-yìg-i.  
 father jaguar kill-EVD.SCND.PAST-3MSG  
 ‘Father killed a jaguar.’<sup>3</sup>

<sup>1</sup>In the rest of the paper I will refer to evidential morphosyntax and semantic-pragmatic evidentiality interpretations by the lowercase Greek letter Eta, clarifying where necessary.

<sup>2</sup>This is, of course, up to interpretation. See Stenzel (2004) for an analysis of Desano and Wanano evidentials as modals (Desano and Wanano are both Eastern Tukanaon languages of Northwest Amazon; closely related to Tuyuca). See also Kaye (1970), which might be the first ‘generative’ analysis of evidentials, and treats Desano evidentials as modals.

<sup>3</sup>Abbreviations for all glosses: EVD = evidential; SCND = secondhand; PST = past tense; DES = desiderative; Q = question; NVIS = nonvisual evidential; PRS = present tense; NEG = negative; 1/2 = first or second person (exclusive).

- (2) Basá-diga-**gari**?  
sing-DES-Q.NVIS.PRES  
‘Do (you) want to sing?’
- (3) Basá-diga-ri-**g**-a.  
sing-DES-NEG-EVD.NVIS.PRES-1/2  
‘(I) do not want to sing’

Generally, in the literature, evidentiality has been analyzed as belonging to one of the functional categories of Aspect, Mood, or Modality, depending on the particular language and accompanying analysis. Others believe it is a grammatical category itself, though this does not entail that ‘evidence’ is a grammatical primitive. Cross-linguistically, then, there appears to be no consensus. I propose to take each analysis at its word and not dispute the variation we find. This follows from the assumption that evidentiality interpretations are to be found in ALL languages and the particular mechanisms responsible for such interpretations are parametrized, namely along projections of either *v*P (as Aspect), or CP (as Mood or epistemic Modality). These functional projections all orbit around the Tense domain in a Cinque (1999) like hierarchy. This makes sense given that TP is not typically understood to be a phase, but *v*P and CP are—that is, morphosyntactic parametrization of the  $\eta$  interpretation can be tied to phase transfer domains. It also makes sense if the featural composition of  $\eta$  includes a structural similarity to Tense as implied in section 2.1

The assumptions I am making here are consistent with something along the lines of the Evidential Domain Hypothesis (EDH) proposed by Blain & Dèchaine (2007).<sup>4</sup>

- (4) **Evidential Domain Hypothesis (EDH)**  
Evidentiality can be interpreted in different phase domains (*v*\*P, CP) cross-linguistically

## 2.1 Feature composition of $\eta$

Functional analysis of evidentiality has been criticized by formalist approaches for not providing a rigorous detailed model of the phenomenon (for example, von Stechow & Matthews (2008: 33)). As a step in this direction I assume that  $\eta$  is composed of some set of finite features, which, together, can operate as a single entity. There are two ways to state it: (i) a syntactic/lexical item may contain an  $\eta$  feature that needs to be checked in

---

<sup>4</sup>Something of a sidenote here: in reference to Anderson’s (1999) comment that many hypotheses in generative grammar are not based on empirical data, and consequently are nothing more than (at best rationally well-grounded) methodological approaches, the EDH appears to be a hypothesis in the normal sense of the term: it was conceived from observations based on empirical data and builds into its definition the (potential) grounds on which it can be proved wrong. Additionally, the hypothesis itself appears to have been proposed independently by different scholars. I myself have only recently read Blain and Dèchaine’s paper but have for the last two years been working on something similar to their observation, see Bowles (2007), Bowles (to appear: footnotes 2, 8) and Bowles (2008: 95).

a probe-goal relation (or some kind of dependency) for semantic-pragmatic interpretation, or, (ii)  $\eta$  can be a head,  $\eta^0$ , that projects to an  $\eta P$  level and licenses certain complements based on the feature composition in  $\eta^0$ . Here (i) rests on lexical entries, and (ii) rests on a predefined structural hierarchy.

Take, for example, the lexical verb *see*. According to (i) we might say that  $see_n$  has a finite number of lexicalized interpretations, one of them being the evidential type,  $see_\eta$ , that means something like “having visual sensory information.” According to (ii), *see* acquires its evidential interpretation by occupying a syntactic position, i.e., the head position of the functional category  $\eta P$ . Independent of choosing (i) or (ii), evidential interpretations can be empirically verified by looking at either the syntactic complements licensed by the item or, in the case of evidential adverbs, the syntactic position. To summarize, I propose a finite set of compositional features that belong to a position in a functional hierarchy and/or a lexical item: these can be seen in Table (1).

Table 1: Feature Composition for  $\eta$ 

ANAPHORIC	SITUATION	SEMANTICS	PRAGMATICS
[+ speaker], [+ deictic sphere]	Reference, Evaluative, Discourse	world = w context = c time = t modal base = B	agent = A quality threshold = $c_\tau$ proposition = p common ground = CG $\mu_c(\mathbf{ev}) = P_c(\varphi_{\mathbf{ev}})$

My starting point for the features in Table (1) are those in [Speas \(2004, 2007\)](#), who proposes a model of featural agreement between modal base and evidential; I also draw heavily from the (probabilistic) pragmatic analysis of ?.

I will not discuss in detail all the elements of Table (1). I would like to focus on the last line in the PRAGMATICS column that shows the mapping function from ? (? : 10): “ $\mu$  maps context-morpheme pairs to probabilities”  $P$  of true propositions in the situation — this is based on the evidential evidence  $\varphi_{\mathbf{ev}}$  (where  $\varphi$  is the true proposition  $p$  and  $\mathbf{ev}$  = evidential). The  $\mu_c$  function is a good candidate for a parameterized feature of evidentiality systems: languages such as English do not have it (or it is dramatically reduced in *Force*), and languages such as Tuyuca do have it.

I hypothesize that an illocutionary  $\eta Force$  position (similar to Declarative and Interrogative Force and discussed below) could be located in three places: (i) within the content of the  $\eta$ -morpheme, which the  $\mu_c$  function maps to a probability — see Figure (??); (ii) in a specifier position of  $ForceP$ , namely in the *spec* of Declarative Force or Interrogative Force both of which are in the CP domain — see Figure (??); or (iii) as a head with a maximal projection,  $\eta^{\max}$  or  $\eta P$  — see Figures (??), (??), and (??). Intuitively,  $\eta Force$  works similar to the way that interrogative or declarative work: it can contribute to the content of lexical items, be in a specifier position in which operator-variable binding applies, or base-generate certain

lexical items. Two examples showing tree structures for the interaction of present tense and a visual  $\eta$ -interpretation in the syntax can be seen in Figure (??) for Tuyuca-like languages and Figure (??) for English-like languages.

The structure in Figure (??)<sup>5</sup> assumes a distinct  $\eta$  category with bare phrase X-bar structure. It is also based on empirical data that show the fusion between tense and evidential inflectional morphology in languages such as Tuyuca (see Bowles to appear for a basic typology of fused tense-evidentials and an account of fusion between  $T^0$  and  $\eta^0$ ). But a distinct category does not have to be assumed and the features that compose  $\eta$ -interpretations could conceivably be distributed across the domains of  $C^0$ ,  $T^0$ , and/or  $v^0$ . Needless to say, given the variation in morphosyntactic evidential systems — and variation in evidentiality in languages without  $\eta$ -morphs — the models in Figures (??) and (??) are not meant to be descriptive of all systems. (Bowles 2010 is an attempt to construct applications of this model to the language Pirahã and, generally, for so-called non-configurational, or discourse configurational, languages).

## 2.2 Evidential Force from a Pragmatic View

? (? : 10-11) state that

Though direct evidence might be reliably better than hearsay evidence, this is not a lexical fact per se, but rather a fact that we derive from general regularities in the world and the context of utterance, and thus it is conceivable that things could be reversed in some situations.

Sentential propositions could possibly show a relative ranking of evidential force ( $\eta Force$ ), which is dependent on the context of unique situations. I define  $\eta Force$  as equivalent with the lower bound (quality threshold) of subjective probability distribution for *Lewisian Quality* defined in ? (? : 7-9).

$$(5) \quad \eta Force = [C_{A,c}(p) \geq c_\tau] \approx .98 \text{ (in the interval } [0,1])$$

Where  $C$  is for credence and an agent  $A$ , in any context  $c$  that has a threshold  $c_\tau \in [0, 1]$ , can assert proposition  $p$ . What might be called a ‘realistic’ constraint on belief attribution keeps a person’s belief below absolute (i.e., below 1 in the numerical interval  $[0,1]$ ; 100% epistemic certainty seems too stringent).

Notice that in the ? (? : 11) model that  $c_\tau$  becomes  $\mu_c(\mathbf{ev})$  before a speech act is performed (their (20.ii)). The concept of  $\eta Force$  captures this for languages like Tuyuca and English: in the former it is realized through the morphology of the evidential system while also functioning as

---

<sup>5</sup>For clarity, I only show a subset of the proposed features in the trees. Also note that, following Speas’ work, I am using a neo-Reichenbachian tense for describing present tense: R,S,E = Reference, Speech, Event. Note also, that the situation features for evidentials, proposed in Speas’ work, is defined similarly to tense.

an operator, in the latter it may be phonetically null (or contribute to a lexical item such as an evidential adverb) but still functions as a discourse operator.

For example, the English sentences in (6) provide a unique situation where two participants are at a concert and direct visual evidence seems to be outranked by indirect auditory. (A related example of evidence re-ranking is found in Kratzer's (2008) 'bear sighting' example; though Kratzer does not explicitly discuss evidentiality).

- (6) a. I see<sub>VISUAL</sub> the band playing.  
       b. I see<sub>VISUAL</sub> the band playing *Hard Rain*.  
       c. I hear<sub>AUDITORY</sub> the band playing *Hard Rain*.

In (6) one might argue that it is conventional to presuppose that visually seeing a band play does not provide the appropriate truth-conditional context in which one can assert or propose to know what song is being played. Credence for the agent in this context has a certain quality threshold value, based on the probability that their proposition is true dependent on the evidence. For (6a) such a threshold is high and the credence of the proposition is likely equal to the threshold — given the proper context. Now, compare (6b) and (6c); it is likely that the evidence required to make the proposition "The band is playing *Hard Rain*" true, depends on what is presupposed to count as verification for knowing that a specific song is being played. In short, hearing a song would generally hold greater satisfaction for credence than seeing the band playing their instruments. For a more dramatic example where evidentiality types would have to be re-ranked, think of providing evidence for the proposition "The Earth orbits the Sun." In such a case, empirically verified observation — based on normal visual evidence from the Earth — has historically given the wrong result. One needs to rank inference higher than visual evidence in order to get the right result; and this is the case historically in many scientific investigations.

Visual evidence is usually the strongest type of direct evidence available for speakers. But this is not a universal property, as the sentences in (6) suggest. The evidentiary basis of evidentiality is often relativized to a situation or context, which itself may depend on certain conversational or conventional implicatures. A model of pragmatic informational strategies, then, *must* incorporate some ability to relativize and rank types of evidentiality within certain contexts and situations.<sup>6</sup> All in all, pragmatics needs to be able to relativize and (re-)rank speaker sources of evidence for propositions.

---

<sup>6</sup>The question of whether or not *Lewisian Quality*, or for that matter *Bayesian Confidence Interval*, is part of natural language grammar is another question. I am inclined to respond by saying that at this point in the history of linguistic science it does not really matter — what matters is that we find models that work and then ask questions. If pushed to answer: *Lewisian Quality* or *subjective quality distribution* is probably not part of human grammar, but it is part of human cognition. In other words, it is a matter of grammar interfacing with other cognitive domains, and thus, is a really difficult puzzle to solve. The best method for solution is 'keep making progress.'

### 2.2.1 The Syntax of Illocutionary ForceP in English Evidentiality

There are two widely accepted cartographies of discourse level syntax that can be used to account for evidentiality. One is the Mood<sub>Evidential</sub> position in Cinque’s (1999) functional hierarchy, the other is Rizzi’s (1997) Force Phrase, which sits atop three other discourse informational positions in the syntax and is responsible for declarative and interrogative force, among others. It is also widely assumed that both Cinque’s and Rizzi’s mappings should overlap in some way — although the details have not been worked out in any depth. An intuitive move here is to assume that because evidentials can occur as discourse modifiers (or operators) for both declarative and interrogative sentences, then the discourse function of the syntax of evidentiality is probably separate but compatible with Declarative and Interrogative ForceP. There are a few ways to model such an intuition, and I will provide one direction here that is compatible with the pragmatic view of  $\eta$ Force sketched above.

The main idea here is that an  $\eta$ Force feature or an  $\eta$ Force projection is co-relational with the syntactic behavior and/or position of ForceP or Mood<sub>Evidential</sub>. In a language like English, an embedded clause with a semantic parenthetical (e.g., “I see” as in “Jon washed his car, I saw him”) that has an  $\eta$ -interpretation can take the form shown in (7) and Figure (??). The embedded clause is modified by a semantically reduced lexical item *see<sub>KNOW</sub>* that contributes to the  $\eta$ -interpretation. In this kind of structure  $\eta$ Force is located with the *that* item. In sentences of the form in (8), the lexical item *see<sub>VISUAL</sub>* carries  $\eta$ Force; see also Figure (??); Gisborne & Holmes (2007) provide historical and corpus data on the different senses of *see* and *know*.

- (7) [ $\eta$  *see<sub>KNOW</sub>* *that* [<sub>CP/TP/vP</sub>...]] = Indirect Reading
  - a. I see that Jon washes his car
- (8) [ $\eta$  *see<sub>VISUAL</sub>* [<sub>CP/TP/vP</sub>...]] = Direct Reading
  - a. I see Jon washing his car<sup>7</sup>

$\eta$ ForceP is in the topmost CP domain of a minimalist style syntax. van Gelderen (2001: 108) notices that “a finite CP contains either a finite complementizer *that* or a sentence adverb (but not both).” If we make the distinction in terms of an  $\eta$ Force feature, then, either a lexical item (e.g., sentence adverb) will carry  $\eta$ Force or the finite complementizer will, but not both. However, this does not appear to be the case in all instances. As van Gelderen (2001: 114) further observes,

Accordingly, the complementizer *that* is in complementary distribution with speech act adverbs (*frankly* [...]) and evaluative adverbs (*surprisingly* [...]), [which is] expected if both are in ForceP. It is marginally in complementary distribution with the

<sup>7</sup>In all examples where it may be the case that an unpronounced *that* could be inserted, assume that I do not intend such a reading.

evidential ones (*allegedly* [...]). It is not clear to me what that means about the position of the evidential adverb.

- (9) van Gelderen’s (2001: 23) Speech Act, Evaluative, and Evidential Adverb Examples
- a. i. \*I know that frankly she should be concerned
  - ii. \*I know frankly that she should be concerned
  - b. \*I know that surprisingly he left
  - c. ?I know that allegedly he left

The lack of clarity for the distribution of evidential adverbials can be explained by an appeal to the variability of  $\eta Force$ . Comparable sentences to (9c) can be seen in Appendix (5.1): such sentences are found on the internet and show that it is possible for  $Mood_{Evidential}$  adverbs to be embedded under the *that* clause. I argue that this is the same *that* clause hypothesized to be the ForceP that carries  $\eta Force$  in Figure (??). An explanation arising out of these details should be along the lines of syntactic-semantic restrictions on the co-occurrence of some of these items. In other words,  $[\eta Force \text{ that}]$  licenses  $Mood_{Evidential}$  adverbials when such adverbials are directly embedded by the *that* clause. And even though there may be some problems with the judgments given for embedded evidential adverbs, there are attested examples. This brings van Gelderen’s statement about “complementary distribution” into question.

Furthermore, we can see that  $Mood_{SpeechAct}$  adverbs might license  $\eta Force$  in limited cases in the examples given in (10); though, of course, such examples need to be empirically tested.

- (10) Speech Act Modifiers
- a. Obviously/Unfortunately, I see Jon mowing his yard in the nude
    - i. Obviously, I see that Jon mows his yard in the nude
    - ii. ? Unfortunately, I see that evidently Jon mows his yard in the nude
    - iii. ? Unfortunately, evidently, Jon mows his yard in the nude
    - iv. ? Unfortunately, I see that Jon evidently mows his yard in the nude
  - b. ? Frankly/Honestly, I see<sup>8</sup> Jon mowing his yard in the nude
    - i. ? Frankly, I see that Jon mows his yard in the nude
    - ii. ? Honestly, I hear that Jon allegedly mows his yard in the nude
    - iii. ? Honestly, allegedly, Jon mows his yard in the nude
    - iv. ? Honestly, I hear that allegedly Jon mows his yard in the nude

---

<sup>8</sup>Again, this should not be read as allowing an unpronounced *that*.



The examples given so far should be enough to speculate a functional ordering based on Cinque's (1999) hierarchy. But as van Gelderen (2005) demonstrates there is one attested example in the British National Corpus (BNC) that goes against the Cinque ordering, though it conforms to the ordering of Rizzi (1997) with *that* in ForceP and the adverb in TopicP. The BNC example is in (11), along with a sample from my search (examples 12–14) of the string **that frankly** in the Corpus of Contemporary American English (COCA).<sup>9</sup> There are, as expected under Cinque's hierarchy, numerous examples with **frankly that** order. Putting some of these observations together we get (15), which is also shown in Figures (??) and (??).

- (11) ...she has told Paul **that frankly** she's lapping up the attention (BNC KBF 8830)
- (12) And I'm very concerned **that frankly**, this is an abuse, potentially, of the mental health system (COCA #3 2007 SPOK NBC.Today)
- (13) But, you know, there are probably many similar cases, Terry, **that frankly** we know nothing about (COCA #8 2005 SPOK NPR.FreshAir)
- (14) We have a competitive frenzy **that frankly** involves parents more than it involves kids themselves... (COCA #15 2004 MAG Psych-Today)
- (15)  $\text{Mood}_{\text{SpeechAct}} > (\eta)\text{ForceP} > (\text{TopicP}) > \text{Mood}_{\text{Evidential}}$

Further evidence for these structures comes from looking at van Gelderen's (2005) analysis of factives (see also Simons (2007)). In her analysis, strings such as **the fact that** can be located in ForceP: with **the fact** located in *spec,ForceP* and **that** in Force<sup>0</sup>. She notes that there are no attested examples in the BNC of the string **see the fact that**. This could be because evidentiality readings do not seem forthcoming when a DP is the complement, as in [see [*DP*]]. In other words, if the lexical item *see* has an  $\eta$  feature it will not license a DP or NP.<sup>10</sup> This may be the case for the factives discussed by van Gelderen (2005).

I agree with van Gelderen (2005) that examples such as (16) should not be attested, but under my analysis we should expect to see examples such as (17).

- (16) \*I **see the fact that** Jon mows his lawn
- (17) I **see that the fact that** 'Jon mows his lawn' is provable by direct observation is a valid way to proceed

<sup>9</sup>This was done on July 01, 2009; I found 3 more examples to add to van Gelderen's. The COCA contains more than 385 million words from spoken and written sources in fiction, popular magazines, newspapers, and academic texts, and is maintained by Mark Davies at Brigham Young University.

<sup>10</sup>The obvious issue here is the sentence *I see Jon*. Without getting into too much detail, sentences of this type seem more to be making a claim on the existence of Jon, and not so much providing a source of evidence for the proposition that Jon exists. However, this explanation does not really satisfy.

- a. I see that  $it_i$  is a valid way to proceed
- b.  $i$  = The fact that ‘Jon mows his lawn’ is provable by direct observation

It turns out that there are no attested cases of (17) in neither the BNC nor the COCA, but this does not mean that (17) is unacceptable. While no corpora show attested examples, a search of the internet does (see Appendix 5.2).

### 3 Beginning to Formalize an Intuition

Group theory is an algebraic formalization useful for studying the symmetry of relations. The notion of antisymmetry is popular in Minimalist syntax theories and symmetry carries some weight in formal semantics (e.g., the use of lattices and morphisms). I have argued elsewhere that this is a good reason to look towards group theory for some inspiration in dealing with formal models for human languages (see Bowles 2009a)<sup>11</sup>. I will apply this same reasoning here, specifically towards formalizing *context*.<sup>12</sup>

#### 3.1 Symmetry and Contexts

Evidential systems are commonly partitioned into Direct and Indirect: that is, directly based on empirical or sensory context, or indirectly based on lower-ranked empirical data or inference — the latter usually being understood as not having ‘direct’ access to the event or information one is commenting on. Both Direct and Indirect types of evidentiality depend on contextual reference of the source of evidence the proposition is based on. The intuition here is that Direct evidentiality can be formalized by showing that contextual relations between speaker and evidence are symmetrical, i.e., they form a group.<sup>13</sup> This requires a formalization of context for evidential reasoning, what could be called an evidentiary base  $B_\eta$ . The

<sup>11</sup>More to the point, many formal techniques used in NLP that rely on linear algebra, vector spaces, or matrices, can be modeled in group theory. Additionally, algebraic structures such as rings, monoids, or various parts of Boolean algebra can also be related to groups by viewing them as extensions or precursors to groups; probability can also be modeled using group theory. Lastly, group theory is useful in working with theories of grammar; e.g., the theory based on Lambek calculus called pregroup grammar (see, for example, Stabler 2004); see also Dymetman (1998a,b) for various grammatical applications.

<sup>12</sup>I am not alone here. McCarthy (1996: 5) notices that “The variety of potential applications of contexts as objects suggests looking at contexts as mathematics looks at group elements.” However, my treatment is different than McCarthy imagines: he seems to imply only an analogy between context and the way group elements are treated in the mathematics of group theory, with contexts being primitives; while I am trying to see how group theory application can be used for atomizing a context so that contexts can be composed of group relations. See Hirst (2000) for criticism of the general enterprise surrounding McCarthy (1996).

<sup>13</sup>However, this may be too stringent in that it would require a logically closed set of propositions for the whole context. Something like  $\eta Force$ , or quality threshold of subjective probability distribution on belief attributions, ensures that a speaker cannot have 100% epistemic certainty in a context  $c$ . It seems that a closed set of propositions for a context must imply 100% epistemic

$B_\eta$  is intended to be compatible with the modal base  $B$  (or more accurately, the *kernal* of the base,  $B_K$ ; see von Fintel & Gillies (2009) who define *kernal* in terms of direct information). Instead of defining relations over an ordering source, the evidentiary base breaks down what a context is by giving definitions for direct or indirect information relative to (non)symmetrical relations that make up the context.

### 3.1.1 Definition of Group

Group theory presupposes some knowledge of set theory because, technically, any group is also a set. That is, any collection of objects, including a set or collection of sets, can be a group if it meets Definition 1; the following is loosely based on Rosen (1995) and Milne (2008).<sup>14</sup>

**Definition 1.** GROUP:

*G is a group iff G, under a law of composition, meets Criterion 1–4.*

**Law of Composition.** (informal):

Any two possible elements,  $a$ ,  $b$ , may be combined in any possible way, where ‘ $\circ$ ’ is any combination, iff  $a \circ b$  equals a set  $S$ .

**Criterion 1.** *Closure*:

For all  $a, b : a, b \in G$ , then  $ab, ba \in G$ .

**Criterion 2.** *Associativity*:

For all  $a, b, c : a, b, c \in G$ , then  $a(bc) = (ab)c$ .

**Criterion 3.** *Existence of Identity*:

$G$  contains identity element  $e : ae = ea = a$ .

**Criterion 4.** *Existence of Inverses*:

For every  $a \in G$ , then  $a^{-1} \in G : aa^{-1} = a^{-1}a = e$ .

A “law of composition” is broadly defined as any procedure that combines any two elements in any way (i.e., a binary operation). This can include the combination of functions  $f(n)$ , relations  $R$ , integer values in addition or multiplication ( $t + y$ ,  $t \times y$ ), or the combination of collections of objects that are subsets of a set,  $s \in S$ . Compositions also include permutations, rotations, and translations by displacement of a certain distance along a certain line. (For more on group theory, see Chapter 10 in Partee et al. (1990), section 39 in Kleene (2002 [1967]); also Dornhoff & Hohn (1978) or any abstract modern algebra book). The genius of GT is that

---

certainty — this is not realistic for a ‘whole’ context: it suggests a speaker knows everything there is to know about a context, or at least knows all the propositions relevant for a context, which is not realistic. This further suggests that a partition for contexts along *local* and *global* lines is useful. A speaker might be allowed a closed set of propositions for a local context — or might be said to have a temporary epistemic certainty of 100% as long as ‘temporary’ is tied to the localness of the context. That is, the local context will be updated, and so will subjective probabilities and epistemic certainty.

<sup>14</sup>It should be noted that definitions can vary slightly in the literature on Group Theory. Milne (2008) does not explicitly state the *closure* criterion, but does make explicit reference to the fact that a group  $G$  is a set. He also refers to the ‘neutral element’ where Rosen uses ‘Identity.’

it is so broadly defined that it can describe a wide variety of concrete or abstract objects, but is rigorous enough to derive very precise relations between those objects. For the purposes here, I use it to formalize a notion of *context* that is compatible with a modal base, and specifically, the *kernal* of a modal base,  $B_K$ .

## 4 Conclusion

I introduced a general sketch of the formalization of some of the syntax, semantics, and pragmatics of evidentiality in English. I used a minimalist style syntax and a probabilistically styled pragmatics and started to extend this model into an algebraic group theory definition of the notion *evidentiary base*; the latter is intended to be fully compatible with a modal base.

Generally, I presented a first step at integrating some of the formal proposals for the natural language phenomenon of evidentiality through combining insights from typological work on grammaticalized evidentials, corpus and historical studies on evidential lexical items in English, and theoretical arguments for hierarchical relations between functional items in information structure.

## 5 Appendix

### 5.1 Some Attested Examples of $\eta$ Force Embedded Evidential Adverbs

Interestingly, none of the corpora I searched had examples, but internet searches did get some results.<sup>15</sup>

#### (18) LinguaLex Examples

- a. On the simple mouse\_test demo **I see that, evidently**, IE8 still doesn't properly support text nodes<sup>16</sup>
- b. For, consulting Omron's own web site, **I see that evidently** they are (as of this writing) in the process of replacing...<sup>17</sup>
- c. **I see that obviously** when you're younger you're not going to be able to do that much by yourself...
- d. Well **I see that obviously** she believes the guy because...
- e. **I hear that...allegedly...** a few other animals in the neighborhood are missing. So sad and disgusting.<sup>18</sup>

<sup>15</sup>The corpora include: Corpus of Contemporary American Speech, British National Corpus, VOICE *online*. The internet searches were done by looking for full strings of the boldfaced material in the Google search engine LinguaLex, which was designed by the author to search through English language sensitive material; more empirical data is needed.

<sup>16</sup>From <http://cross-browser.com/forums/viewtopic.php?id=515>.

<sup>17</sup>From <http://www.epinions.com>.

<sup>18</sup>This was a written text and it appears that the ellipsis marks are not intended to delete any material.

- f. It may be just a rumor, but **I hear that allegedly** The Public can't afford to pay the bills for the electricity...

(19) Bing Examples

- a. **I hear that allegedly** there is a separate room in the Family and Adult video store...
- b. **I see that allegedly** there are no responses at this time to that particular posting...

These examples are clearly problematic, but not so much so that they are not suggestive of further directions for the cartography of the hypothetical  $\eta$ Force and the kinds of complements it licenses; particularly in terms of its interaction with TopicP elements and what appears to be the licensing of a parenthetical evidential adverb within the intonational pause after the *that* element. Also, the examples (18a) and (18b) suggest some interesting issues: in the case of (18a) the first clause *On the simple mouse\_test demo* seems to be focused (new) material in a FocusP position; in the second case of (18b) the position of *I see that evidently* comes after what appears to be the Focused content (*consulting Omron's own web site*), which is embedded under the *For* item — which itself is probably FinP.

## 5.2 'I see that the fact that...'

(20) LinguaLex Examples

- a. **I see that the fact that** extensions are not enabled by default is confusing and will try to make this more clear in final 1.1 version
- b. **I see that the fact that** human beings have a free will and, unlike mindless animals, can exercise moral and behavioral choice was also lost on you...<sup>19</sup>
- c. So **I see that the fact that** it happened after the installation of the psu must be coincidental<sup>20</sup>

<sup>19</sup>From <http://www.dakotavoices.com/2009/05/>.

<sup>20</sup>From <http://h30434.www3.hp.com/psg/>.

## References

- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Anderson, Stephen R. 1999. A formalist's reading of some functionalist work in syntax. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan & K. Wheatley (eds.) *Functionalism and formalism in linguistics*, vol. 1, 111–135. Amsterdam: John Benjamins.
- Barnes, Janet. 1984. Evidentials in the Tuyuca verb. *International Journal of American Linguistics* 50: 255–271.
- Barnes, Janet. 1994. Tuyuca. In Peter Kahrel & Renè van den Berg (eds.) *Typological studies in negation*, 325–342. Amsterdam: John Benjamins.
- Barnes, Janet. 1996. Autosegments with three-way lexical contrasts in Tuyuca. *International Journal of American Linguistics* 62: 31–58.
- Blain, Eleanor M. & Ros-Marie Déchaine. 2007. Evidential types: Evidence from Cree dialects. *International Journal of American Linguistics* 73(3): 257–291.
- Bowles, Joshua. 2007. Phi and Eta interactions in Tuyuca. Ms., University of Utah.
- Bowles, Joshua. 2008. *Agreement in Tuyuca*. Master's thesis, University of Utah.
- Bowles, Joshua. 2009a. Merge, symmetry, and group theory $\alpha$ . Ms., University of Utah and Utah Valley University.
- Bowles, Joshua. 2009b. Tuyuca data set. Ms., University of Utah and Utah Valley University.
- Bowles, Joshua. 2010. Merge, groups, (anti)symmetry, and evidentiality: Hierarchical relations in information structure $\alpha$ . Under review for presentation at the WCCFL 28.
- Bowles, Joshua. to appear. The morphosyntax of fused tense-evidentials: An initial model. *Linguistic Student Organization Working Papers in Linguistics* 8. URL <http://ling.wisc.edu/lso/wpl-main.html>.
- Campbell, Lyle. 1991. Some grammaticalization changes in Estonian and their implications. In Elizabeth C. Traugott & Bernd Heine (eds.) *Approaches to Grammaticalization*, vol. 1, 285–299. Amsterdam: John Benjamins.
- Chung, Kyung-Sook. 2005. *Space in Tense: The interaction of tense, aspect, evidentiality, and speech act in Korean*. Ph.D. thesis, Simon Fraser University.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads*. Oxford: Oxford University Press.
- Dornhoff, Larry L. & Franz E. Hohn. 1978. *Applied modern algebra*. New York: Macmillan Publishing Co., Inc.
- Dymetman, Marc. 1998a. Group theory and computational linguistics. *Journal of Logic, Language, and Information* 7: 461–497.
- Dymetman, Marc. 1998b. Group theory and linguistic processing. In *COLING-ACL*, 348–352.
- Faller, Martina. 2002. *Semantics and pragmatics of evidentials in Cuzco Quechua*. Ph.D. thesis, Simon Fraser University.

- Garrett, Edward. 2001. *Evidentiality and assertion in Tibetan*. Ph.D. thesis, University of California, Los Angeles.
- Gisborne, Nikolas & Jasper Holmes. 2007. A history of English evidential verbs of appearance. *English Language and Linguistics* 11(1): 1–29.
- Harris, Alice & Lyle Campbell. 1995. *Historical Syntax in a Cross-linguistic Perspective*. Cambridge: Cambridge University Press.
- Hirst, Graeme. 2000. Context as a spurious concept. In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics*, 273–287. URL <http://ftp.cs.toronto.edu/pub/gh/Hirst-CICLing-2000.pdf>.
- Johnson, Lars & Bo Utas (eds.) . 2000. *Evidentials: Turkic, Iranian and neighboring languages*. Berlin: Mouton de Gruyter.
- Karn, Gloria Jean. 1976. *Tuyuca plot structure: A pilot study*. Master's thesis, University of Texas, Arlington.
- Kaye, Jonathan D. 1970. *The Desano verb: problems in semantics, syntax and phonology*. Ph.D. thesis, Department of Philosophy, Columbia University.
- Kleene, Stephen. 2002 [1967]. *Mathematical logic*. New York: Dover.
- Kratzer, Angelika. 2008. Situations in natural languages semantics. In Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Stanford University: Metaphysics Research Lab, CSLI. URL <http://plato.stanford.edu/entries/situations-semantics/>.
- McCarthy, John. 1996. A logical AI approach to context. URL <http://www-formal.stanford.edu/jmc/>. Unpublished note, Stanford University.
- Milne, James S. 2008. Group theory. Lecture Notes, University of Michigan.
- Partee, Barbara, Alice ter Meulen & Robert E. Wall. 1990. *Mathematical methods in linguistics*. Dordrecht: Kluwer Academic Publishers, 2nd edn.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.) *Elements of grammar*, 281–338. Dordrecht: Kluwer Publications.
- Rosen, Joe. 1995. *Symmetry in science: An introduction to the general theory*. New York: Springer-Verlag.
- Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117(6): 1034–1056.
- Speas, Margaret. 2004. Evidential paradigms, world variables and person agreement features. *Italian Journal of Linguistics* 16(1): 104–135.
- Speas, Margaret. 2007. Evidentials as generalized functional heads. Ms., University of Massachusetts, Amherst.
- Stabler, Edward. 2004. Tupled pregroup grammars. URL <http://www.linguistics.ucla.edu/people/stabler/epspub.htm>. Draft., University California Los Angeles.
- Stenzel, Kristine. 2004. *A reference grammar of Wanano*. Ph.D. thesis, University of Colorado, Boulder.
- van Gelderen, Elly. 2001. The force of ForceP. *Southwest Journal of Linguistics* 20(2): 107–120.
- van Gelderen, Elly. 2005. The CP and split CP cross-linguistically. *WORD*

55(3): 369–403.

- von Fintel, Kai & Anthony S. Gillies. 2009. *Must*  $\lambda$ dots stay  $\lambda$ dots strong!  
URL <http://mit.edu/fintel/fintel-gillies-2009-mss.pdf>. Ms., draft version of July 17, 2009, MIT, and Rutgers University.
- von Fintel, Kai & Lisa Matthewson. 2008. Universals in semantics. *The Linguistic Review* 25(1-2): 139–201.