THE LEARNABILITY OF SEMANTIC DISTINCTIONS: THE CASE OF EVIDENTIALITY

by

Dionysia Saratsli

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics

Spring, 2022

THE LEARNABILITY OF SEMANTIC DISTINCTIONS: THE CASE OF EVIDENTIALITY

by

Dionysia Saratsli

Approved: ____

Benjamin Bruening, PhD.
Chair of the Department of Linguistics and Cognitive Science

Approved: ____

John Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: ____

Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Anna Papafragou, Ph.D.
Professor in charge of the dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Zhenghan Qi, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Satoshi Tomioka, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Helene Intraub, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

I would like to start by thanking my advisor, Anna Papafragou, as her expertise and continuous support and mentoring have been the cornerstone of this work. Anna throughout the years "molded" me into the scientist and writer that I am today. Being the first one in my family to get into college and then into graduate school, Anna has been an inspiring role model of a successful woman in science, and I am truly grateful that I had the opportunity to work with her. I would also like to thank my committee members, Zhenghan Qi, Satoshi Tomioka and Helene Intraub for their support, genuine interest in my work and all the insightful comments and questions. I would like to add special thanks to one of my committee members, Zhenghan Qi, for allowing me the opportunity to work with her in an additional project. For the purposes of this project, Zhenghan welcomed me in her lab and has been tirelessly mentoring and helping me throughout this work. Through her training and expertise, I have learned a lot and I have been continuously inspired by her passion for research and for that I am truly thankful.

I would like to extend my gratitude to present and past members of the Language and Cognition lab for their feedback, discussions and camaraderie. Special thanks go to Myrto Grigoroglou for her friendship and advice. I would also like to thank Ariel Mathis for being an amazing lab manager, timely and tirelessly setting up and running all the different studies of this work.

Special thanks to Leventis Foundation for the Fellowship that supported my graduate studies, the Office of Graduate and Professional Education for a Fellowship that funded participant e recruitment and the Department of Linguistics and Cognitive Science for a series of generous Summer Fellowships that supported data collection and manuscript writing.

I would also like to thank all my friends, here in the US and back home, for their continuous love and support throughout the years. Specifically, I would like to thank my amazing Newark-based friends (Katarina Smiljanec, Kim Milla, Dimitris Mouris, Eirini Vangeli, Bjorn Hasa, Tina Doika, Nefeli Kamarinopoulou, Monty Fales, Nick Chobani just to name a few) for listening to my troubles and anxiety, sharing drinks and life adventures. I also want to thank my University of Delaware's Crossfit family for being supportive, teaching me how to always find a way to get better and providing a fun outlet for me. I also want to thank Alena Aksenova, for making the Google Internship and summer 2019 in general, one of the most fun and meaningful experiences – it was my pleasure sharing this intellectually challenging and eye-opening summer with her. I also want to thank all my life-long friends back home in Greece for believing in me and supporting me in this dream to pursue my graduate career here in the US.

I want to express my special thanks and my gratitude to Christodouli Chatzimousi for being my friend for the last 15 (!) years – she was always there for me, pushing me to believe in myself and fight for my dreams despite the adversities but also being there for all the mischievous school moments, endless jokes and laughs and for all those life memories throughout all those years.

I also want to express my gratitude and extend my extra-special thank you to Konstantinos (or Kost, as my wedding band says) – for being my co-pilot in this life journey. One day, we gathered almost all our belongings, and with four suitcases and our fluffy friend Jeff, we chased our dreams together. Thank you for your support through all these endless hours of work and all-nighters, for your love, for the crazy, spontaneous trips, and for the amazing food you cook. Thank you for being my best friend and partner-in-crime.

Above all, I want to thank my parents for their unconditional love and for always being my side, supporting my every dream and decision even if that took me so many many miles away from them. They were the ones that instilled in me the sense of trying your best, without worrying about whether you are the best or whether you will fail. Mom, Dad, you have been the biggest inspiration, support and role models in my life and for that, I have no words to express my gratitude. This is dedicated to you.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

It is often assumed that cross-linguistically more prevalent distinctions are easier to learn potentially due to their conceptual naturalness. Prior work supports this hypothesis in phonology, morphology and syntax but has not addressed semantics. This work aims to unravel the potential factors that contribute to the learnability and the cross-linguistic prevalence of certain semantic distinctions over others. Within the semantic domain, *evidentiality* offers an excellent test case for learnability patterns due to the cross-linguistic frequency of certain evidential systems along with the fact that such distinctions are not grammaticalized in English and can be taught to English-speaking adults within an Artificial Language Learning paradigm without native language interference.

Our first aim is to test to test whether conceptual naturalness can account for ease of learning and cross-linguistic frequency patterns of evidential systems. When exposed to these miniature evidential morphological systems, adult speakers of English learned the typologically most prevalent system (marking indirect, reportative information) better compared to less-attested systems (Experiments 1-2). Similar patterns were observed when non-linguistic symbols were used to encode evidential distinctions (Experiment 3).

Our previous experiments have shown that learners preferentially encode reportatives compared to other sources but left open the nature of the asymmetry because of design-specific issues. Our second aim is to adjudicate between a pragmatic bias and category partition issues as an explanation for the learnability of evidential distinctions.

In Experiments 4-6, we directly compared two simple evidential systems, each marking only a single source (visual vs. reportative) and leaving the other source unmarked. Similarly to our previous findings, participants learned more easily to mark reportative information sources over visual/direct sources of information. Our results provide support for a pragmatic bias that shapes both the cross-linguistic frequency and the learnability of evidential semantic distinctions.

Our last aim is to test whether learnability patterns change when the same information source meanings are mapped on different linguistic and non-linguistic forms. Developmentally, evidential meanings seem to follow a protracted trajectory. According to a prominent hypothesis, this difficulty stems from the complexity of the underlying concepts. On an alternative proposal, the difficulty often lies in the mapping between linguistic expressions and concepts, even if the concepts themselves are available. In Experiments 7-8, we offer a novel argument for the role of mapping factors in acquiring a evidential meanings, specifically pointing to the importance of the syntactic and pragmatic factors that lead to the form-meaning association.

PREFACE

This dissertation includes the dissertation author's previously published work that was completed at the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics and Cognitive Science.

Chapter 2 appeared as Saratsli, D., Bartell, S. & Papafragou, A. (2020). Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality. *Cognition*, 197. https://doi.org/10.1016/j.cognition.2020.104194 .

A modified version of Chapter 3 will be submitted as a revise and resubmit journal submission as Saratsli, D.., & Papafragou, A. (2022). Pragmatic Effects on Semantic Learnability: Insights from Evidentiality. to the *Journal of Memory and Language*.

The dissertation author is the first and corresponding author on both publications. For both publications, the dissertation author designed the experiments, collected the data, performed the statistical analyses and wrote the manuscripts.

Chapter 1

INTRODUCTION

1.1. Cross-linguistic frequency and learnability of semantic distinctions

It is often assumed in the literature that cross-linguistically more frequent distinctions are easier to learn than less frequent ones (Jacobson, 1971; Rosch, 1972; Clark, 1976; Pinker, 1984; Slobin, 1985; Bowerman, 1993). This idea has been captured effectively by Gentner and Bowerman's (2009) Typological Prevalence Hypothesis (TPH): "All else being equal, within a given domain, the more frequently a given way of categorizing is found in the languages of the world, the more natural it is for human cognizers, hence the easier it will be for children to learn" (p. 467). Even though this idea is fundamental for how human beings learn to encode meaning in language, empirical tests of this hypothesis within the semantic domain have so far been rare.

Gentner and Bowerman (2009) themselves tested this hypothesis by comparing how English-speaking and Dutch-speaking children acquire their native language's spatial support prepositions. English encodes spatial support with the single preposition *on*, while in Dutch three different prepositions (*op, aan, om*) partition the same semantic space. Because the English option is typologically more common, the TPH predicts that the English system should be more easily learned than the Dutch system. Gentner and Bowerman elicited spatial prepositions from young learners of the two languages, and the results supported this prediction. However, the slower acquisition rate could be due to the higher number of prepositions (and/or the lower frequency of each individual preposition) in Dutch compared to English rather than the specific semantic distinctions encoded within each system. This fact complicates the interpretation of Gentner and Bowerman's results and hence the evidence in favor of the TPH. Similar issues are bound to arise for all tests of the TPH that involve comparisons of how young learners acquire semantic systems of variable complexity cross-linguistically.

A growing research strand that explores different biases in learning and that could potentially offer valuable evidence on the factors that motivate recurring distinctions across languages is that of Artificial Language Learning. Studies that utilize an Artificial Language Learning paradigm typically require participants to learn different versions of a target miniature language that differ minimally from each other in terms of a grammatical or lexical feature (see Folia, Uddén, de Vries, Forkstam, & Petersson, 2010 for a review). During an initial learning phase, participants are exposed to the grammar/lexicon of the artificial language. The learning phase is followed by a test phase in which the extent to which participants learned the linguistic target is assessed. This paradigm offers a unique opportunity to explore the participants' learning process in relation to a specific linguistic feature of interest (Fedzechkina, Newport & Jaeger, 2016). By having participants learn minimally different versions of the same artificial language, one can bypass the role of frequency or complexity in the learnability of actual systems in individual languages. Moreover, it is possible to test adults on the target artificial language, thereby eliminating the possibility that any learnability patterns observed could be due to cognitive-developmental limitations in the learners themselves.

This experimental paradigm has been used extensively in the study of morpho-syntactic (Smith & Tsimpli, 1991; Christiansen, 2000; Newport & Aslin, 2004; Hudson Kam & Newport, 2005, 2009; Thompson & Newport, 2007; Wonnacott, Newport, & Tanenhaus, 2008; Tily, Frank, & Jaeger, 2011; Merkx, Rastle, & Davis, 2011; Culbertson & Smolensky, 2012; Fedzechkina, Jaeger, & Newport, 2012; Tabullo, Arismendi, Wainselboim, Primero, Vernis, Segura, Zanutto, & Yorio. 2012; Culbertson, Franck, Braquet, Barrera Navarro, & Arnon., 2020) and phonological phenomena (Seidl & Buckley, 2005; Wilson, 2006; Finley & Badecker, 2009), and has only recently been extended to semantics (Xu, Dowman & Griffiths, 2013; Carstensen, Xu, Smith & Regier, 2015; Kemp, Xu & Regier, 2018; Maldonado & Culbertson, 2020).

Concerning meaning distinctions, recent work discusses this constrained cross-linguistic variation of meanings within specific semantic domains in relation to a drive for communication efficiency (Kemp, Xu & Regier, 2018). This idea of efficiency highlights the fact that language, as a vessel for communication, needs to be both maximally informative but also simple: the intended message should contain all the essential information needed to fully understand what is being communicated but this needs to happen using minimal cognitive resources (Grice, 1989; Rosch, 1999; Regier, Kemp, & Kay, 2015; Regier, Carstensen, & Kemp, 2016; Gibson, et. al., 2019; see also: Steinert-Threskeld and Szymanik, 2020 and Carr, Smith, Culbertson and Kirby, 2020 for a more computational methodology). This proposal resonates with developmental work on how informativeness helps children understand and produce language, tracking relevant and pragmatically notable aspects in their input (Frank & Goodman, 2014; Bannard et al., 2017). However, there is no extensive empirical work specifically exploring the learnability patterns along with the potential factors that contribute to the prevalence of certain meaning distinctions over others across different languages.

1.2. Evidentiality across languages

Evidentiality refers to the way that language marks the speaker's source of information, for instance, whether the speaker had direct perceptual access to an event, inferred what happened based on some clues, or was told what happened by someone else. Languages differ in the way they encode evidentiality (Chafe & Nichols, 1986; Willet, 1988; Kratzer, 1991; Cinque, 1999; Johanson & Utas, 2000; Ifantidou, 2001; Mushin, 2001; Delancey, 2002; Aikhenvald & Dixon, 2003, 2014; McCready, 2008; McCready & Ogata, 2007; Aikhenvald, 2004, 2014, 2018). English and other languages make use of lexical means such as verbs (e.g., *see, hear, infer*) or adverbs (e.g., *allegedly*) to mark information sources. About a quarter of the world's languages, however, use grammatical morphemes to indicate information sources (Izvorski, 1998; Garrett, 2000; Faller, 2012, 2014; Matthewson, 2012; Speas, 2004, 2018). For example, Wanka Quechua has separate grammatical morphemes for three types of information source (Aikhenvald, 2004): -*mi* in (1) marks the speaker's visual experience, -*chr*- in (2) marks an inference drawn by the speaker, and –*shi* in (3) marks another person's report about what happened.

(1) Chay-chruu-**mi** achka wamla-pis walashr-pis alma-ku-lkaa-ña.
this-LOC-**DIR.EV** many girl-TOO boy-TOO bathe-REEL-IMPF.PL-NARR.PAST.
'Many girls and boys were swimming' (I saw them).

(2) Daañu pawa-shra-si ka-ya-n-**chr**-ari.
Field finish-PART-EVEN be-IMPF-3-**INFR**-EMPH.
'It (the field) might be completely destroyed' (I infer).

(3) Ancha-p-**shi** wa'a-chi-nki wamla-a-ta.
too.much-GEN-**REP** cry-CAUS-2 girl-1P-ACC.
'You make my daughter cry too much' (they tell me).

Semantically, the broad meaning dimensions that characterize grammaticalized evidentiality across languages involve perception, inference and verbal report (Aikhenvald, 2004, 2014, 2018; de Haan, 2013b; Willett, 1988). Perception mostly includes visual access but also non-visual perception (information acquired through hearing that may also extend to other senses; Aikhenvald, 2018). Inference has several sub-types, including inference from visual premises, general knowledge or other types of reasoning. Verbal report covers cases where the source of the report is left unspecified (as in hearsay), and other cases where a specific entity is quoted. Languages vary in both the complexity of their evidential systems (i.e., the evidential categories they mark and hence the number of distinct evidential morphemes they have) and the way they group semantic distinctions (Aikhenvald, 2004; Speas, 2018). For instance, as discussed in more detail below, very often languages have a single evidential morpheme but leave other sources unmarked. Some languages have only two grammatical morphemes to mark information source, a direct/firsthand morpheme that may encode perceptual sources and an indirect/non-firsthand morpheme that may cover both inference and report (Aikhenvald, 2004). Others encode all three types of information source with dedicated morphemes, as in the examples (1)-(3) above. Yet other languages have more complex 4-way or 5-way evidential systems that subdivide these broad categories into more specialized distinctions (ibid.).

Pragmatically, the use of evidentially marked utterances can give rise to contextual implications about speaker certainty or reliability, even though evidential morphemes do not denote these dimensions as part of their core meaning (Speas, 2018). The pragmatic profile of evidentials is captured by current formal theories, according to which the use of evidentiality introduces a proposal to add the base proposition to the common ground, but whether the addressee accepts this addition depends on how the addressee evaluates the nature of the evidence (Murray, 2011, 2014; cf. AnderBois, 2014). In this sense, linguistic evidentiality connects to the human ability to reason about sources of information (see Papafragou, Li, Choi & Han, 2007; Ünal & Papafragou, 2018, for extensive discussion).

In cognition, direct perceptual experience of an event is regarded as a more reliable source due to its assumed correspondence with reality; indirect sources of information such as inference or reports are often deemed less reliable in the sense that the former may be based on incomplete premises while the latter depends on the informant's reliability (Papafragou et al., 2007; Dancy, 1985; Matsui & Fitneva, 2009; Aikhenvald, 2018; Wiemer, 2018; Koring & De Mulder, 2014). These intuitions emerge early: even at the age of 3, young children know that seeing leads to knowing (Pratt & Bryant, 1990; Pillow, 1989), but identifying the contribution of other sources may be a more protracted development (Ozturk & Papafragou, 2016). Furthermore, children prioritize seeing over inference and hearsay when choosing whether someone else is knowledgeable about an event (Ozturk & Papafragou, 2016), though these preferences are flexible (Brosseau-Liard & Birch, 2011; Fitneva, Lam, & Dunfield, 2013).

Depending on the specific alternatives available within an evidential system in a language, then, the use of an indirect evidential might indicate that the speaker lacks direct evidence for an assertion (Speas, 2018) and might further convey tentativeness, doubt, or distance from the event. For the same reason, when evidence about an event is available from multiple sources, languages exhibit a preference for encoding it with visual evidentials, when possible, while using reportative evidentials is the least preferred option (Aikhenvald, 2018). As Aikhenvald notes, communicating an event that the speaker has experienced "is considered a better choice rather than reporting what they [the speaker] heard from someone else" (2004: 307). This emphasizes "the preference for visual information source – if it is available" and also speaks for the reliability of visual access, other things being equal (Aikhenvald, 2018: 27).

For present purposes, a key fact about grammaticalized evidentiality is that, across languages, there is widespread preference for evidential systems to mark indirect compared to direct access to information (de Haan, 2013a; Aikhenvald, 2003, 2004, 2018). For instance, the World Atlas of Languages lists 161 languages worldwide that grammatically mark only indirect evidence, 71 languages that mark both direct and indirect evidence, and no languages that mark direct evidence alone (de Haan, 2013a). Similarly, according to Aikhenvald (2003, 2004, 2018), many languages have a dedicated reportative morpheme (or, less frequently, an indirect/non-first hand morpheme that can cover inference or reports) but leave other sources unmarked. Systems that mark only direct (visual) access with dedicated grammatical devices but do not grammatically encode other source types are unattested and thus in all likelihood rare (ibid.). As a result, in many languages, "the least formally marked verb in a language with evidentiality tends to acquire a visual, or firsthand reading" (Aikhenvald, 2018, p.16), even though distinguishing such meanings from source-neutral meanings can be difficult). These facts suggest a markedness hierarchy for evidentiality, with direct (visual) evidence being the unmarked case.

Why should direct visual access to information be the least frequently marked morphologically across languages? Previous commentators have observed that "the tendency to mark direct, or visual, or sensory evidentials less than others may reflect the primacy of vision as an information source" (Aikhenvald, 2018, p.16) but the specifics of this connection have remained vague. According to TPH mentioned earlier (Gentner & Bowerman, 2009), this typological frequency indexes the 'naturalness' of the underlying concepts. However, this proposal seem to sit uneasily with the fact that visual perception is known to be a salient, important and natural conceptual category. Visual perception is richly represented in verb meanings across languages (e.g., Viberg, 1984; San Roque et al., 2015; 2018; Majid et al., 2018), and these meanings often get extended to refer to other cognitive processes in ways that are not true of other senses (e.g., in English, *see* can mean ‘realize’; San Roque et al., 2018; Sweetser, 1990). Moreover, the fact that vision gives access to information seems to be available to both seeing and blind individuals (Landau & Gleitman, 1985; Koster-Hale, Bedny & Saxe, 2014).

This dissertation puts forward the proposal that the broad typological facts may be related to the pragmatic implications carried by different sources of information. More specifically, the fact that direct access is generally (even though not always) more reliable compared to less direct sources: put simply, languages tend to grammatically mark potentially less reliable information sources. This perspective coheres with proposals according to which human cognition is equipped with epistemic vigilance so to avoid unreliable sources and the risk of being misinformed (Sperber, Clement, Heintz, Mascaro, Mercier, Origgi & Wilson, 2010).

We know that even young children track a speaker’s trustworthiness and choose to learn things from reliable over unreliable speakers (Sabbagh & Baldwin, 2001; Koenig, & Harris, 2005; Jaswal, & Neely, 2006; Fusaro, Corriveau, & Harris, 2011; Harris, 2012; Koenig, 2012; Mascaro & Sperber, 2009; Jaswal, 2010). However, constantly exercising epistemic vigilance could entail an additional processing cost: the assumption that human communication is presumed to be truthful and informative (Grice, 1989) would be violated, and speakers would need to evaluate not only the actual information they receive but also their interlocutor’s reliability and intentions. One way of providing essential information within a fully cooperative communicative context while maintaining constant epistemic vigilance would be to only mark indirect, potentially unreliable – but not direct perceptual, presumably more reliable - sources of one's experience.

According to this perspective, non-perceptual sources are selectively marked in languages that have a single evidential morpheme because they are *informative*, i.e., they represent a departure from the primacy of perception as an information source (cf. Barnard, Rosen & Matthews, 2017, on a similar notion of informativeness). Furthermore, given that reportative access is the least direct type of information source (since the speaker need have experienced no part of an event), it would be the most likely to be encoded when languages have a single evidential morpheme.

1.3. Outline of the Dissertation

The overall goal of this dissertation is to unravel the potential factors that contribute to the learnability and the cross-linguistic prevalence of certain semantic distinctions over others. Within the semantic domain, *evidentiality* offers an excellent test case for learnability patterns due to the cross-linguistic frequency of certain evidential systems along with the fact that such distinctions are not grammaticalized in English and can be taught to English-speaking adults within an Artificial Language Learning paradigm without native language interference.

Chapter 2 aims to test the predictions set by the Typological Prevalence Hypothesis (TPH) and whether conceptual naturalness can account for ease of learning and cross-linguistic frequency patterns of evidential systems. More specifically, in

Experiments 1 and 2, adult speakers of English whose language does not encode evidentiality grammatically are exposed to miniature evidential systems that marked through a novel morpheme one information source: either the speaker's direct visual experience, speaker's inference process of an event or reportative information. In Experiment 3, participants are asked to learn evidential systems that mark the target information source through a non-linguistic symbol. The goal is to compare participants' learnability patterns for these miniature evidential systems and test whether the most cross-linguistically prevalent system will be the one that will be most easily learned, as predicted by *TPH*.

Building on the findings of Chapter 2, Chapter 3 aims to adjudicate between a pragmatic bias and category partition issues as an explanation for the learnability of evidential distinctions. Towards this aim, adult English speakers were exposed to novel miniature evidential morphological systems that only marked an indirect (reportative) or direct (visual) information source (Experiment 4). In an effort to eliminate any ambiguity or task-related confusion, participants' learnability patterns for the same information sources are also tested after they are given specific cues to the target evidential meanings (Experiment 5). To further test, the pragmatic account for the learnability patterns observed, participants had to learn to encode evidential meanings through visual, not linguistic, means (Experiment 6).

Lastly, Chapter 4 asks whether learnability patterns change when the same information source meanings are mapped on different linguistic and non-linguistic forms. Using an artificial language learning paradigm, we compare adult learners' acquisition of a single evidential meaning expressed by different linguistic forms (a novel verb, morpheme, or adverb) and non-linguistic symbol as a control condition. The expectation is that the learnability of evidential meanings would differ depending on the linguistic and extra-linguistic (pragmatic) properties of the forms that encode these meanings and, correspondingly, the tools that learners use to extract the commonalities within a particular set of events during form-to-meaning mappings.

Chapter 2

THE LEARNABILITY OF EVIDENTIAL PATTERNS: CONCEPTUAL AND PRAGMATIC CONSIDERATIONS

Our first aim was to offer a new test of the predictions set by TPH using an *Artificial Language Learning* (ALL) Paradigm. The Typological Prevalence Hypothesis predicts that a simple evidentiality system that only marks reported information would be more easily learnable compared to a system that only marks visually acquired information (with a system marking only inferred information falling in between). Towards this aim, we used an Artificial Language Learning paradigm to assess the learnability of different evidential morphological systems and test predictions of the TPH (Experiments 1 and 2). Second, we explored whether the learnability patterns for evidential morphology would generalize to non-linguistic symbols to throw light on the nature and origins of the learnability facts (Experiment 3).

Specifically, in our first two experiments, we tested evidential distinctions in an Artificial Language Learning task using adult native speakers of English. Participants were exposed to scenarios in which characters experienced a series of events. Access to the events was manipulated within subjects such that a character might (a) see an event directly, (b) see visual clues and infer that a certain event had occurred, or (c) be informed by an observer about the occurrence of an event. The character described each event in an artificial language containing a novel evidential morpheme (*ga*).

Prior studies with adults indicate that mental state meanings for novel predicates are hard to infer from observing events in the world without specific syntactic support (Gillette, Gleitman, Gleitman, & Lederer, 1999; Papafragou, Cassidy, & Gleitman, 2007). Additionally, the acquisition of evidentiality in children is protracted cross-linguistically (Aksu-Koç, 1988; Lee & Law, 2000; Aksu-Koç & Alici, 2000; Matsui, Miura & McCagg, 2006; Matsui, Yamamoto, & McCagg, 2006; Papafragou et al., 2007; Aksu-Koç, Ögel-Balaban & Alp, 2009; de Villiers, Garfield, Gernet-Girard, Roeper, & Speas, 2009; Winans, Hyams, Rett, & Kalin, 2014; Ozturk & Papafragou, 2016; Uzundag, Taşçı, Küntay, & Aksu-Koç, 2018; Ünal & Papafragou, 2018; Fitneva, 2008, 2018). As a result, we anticipated that acquiring evidential meanings would be challenging. For this reason, we simplified the artificial language that participants were exposed to by preserving the lexicon of English but changing the morphosyntax (the new language lacked determiners and had a Subject-Object-Verb word order). The evidential morpheme appeared verb-finally (*The girl cups stackedga*). The type of access marked by the morpheme was manipulated between subjects such that the morpheme marked a single access type and the other two types received null marking. Thus, depending on the marked meaning, each participant was exposed to a Visual, Inferential, or Reportative evidential system (see Table 1), and had to learn what the morpheme meant. Of interest was whether, as predicted by the TPH, participants would succeed with Reportative systems more than others.

Evidentiality System	**Speaker's Access to Event**		
	Visual (Visual Perception)	Inferential (Inference from Visual Premises)	Reportative (Testimony from Others)
Visual	*ga*	--	--
Inferential	--	*ga*	--
Reportative	--	--	*ga*

Table 1 Structure of evidential systems in Artificial Lanfguage Learning studies (Exps 1 and 2)

In our final experiment, we focused on Artificial Symbol Learning. The design was similar to Experiment 2, but the character's description of the event was in English instead of the artificial language. We replaced the morpheme ga with a non-linguistic symbol (circle) that served as an evidentiality morpheme. If the results of both the morpheme and the symbol learning experiments converge and reveal a preference for marking indirect over direct sources, we will have evidence that both the typological frequency and learnability facts about evidentials are motivated by broader considerations of which distinctions are more 'natural' to encode.

2.1. Experiment 1

2.1.1 Methods

2.1.1.1. Participants

We recruited 111 participants in total, between the ages of 18 and 22 (10 male and 101 female). All participants were undergraduate students at the University of Delaware and were enrolled in an Introductory Psychology course that awarded credit for their participation. They were native speakers of English. Nine participants reported speaking a different language at home but none of these languages included grammaticalized evidentials.

2.1.1.2. Stimuli

The stimuli consisted of 39 scenarios showing a puppet (henceforth, the "Agent") enact an event on the stage of a puppet theater. Another puppet (henceforth, the "Speaker") experienced this event in more or less direct ways (sometimes from the other side of the stage or through an "Informant") and later described it. Each scenario involved a change of state and unfolded over four static pictures presented simultaneously for 20 sec (see Figure 1). There were 3 versions of each scenario depending on the Speaker's access to the event.

(A) In the Visual Access version, the stage curtains remained open and the Speaker could see the event in its entirety. Specifically, the Speaker (e.g., the dog) saw the Agent stand next to some objects (e.g., the gorilla next to the board; panel A1 in Figure 1), manipulate the objects (color the star on the board, panels A2-A3) and stand next to the event aftermath (the gorilla standing next to the colored star, panel A4).

(B) In the Inferential Access version, the Speaker (e.g., the dog) could initially see the Agent and objects, just as in the Visual Access version (e.g., the gorilla and the board with the star, see panel B1). Then the stage curtains closed (panels B2-B3) and later opened to reveal (what presumably was) the aftermath of the event (panel B4; the gorilla next to the colored star). Thus, the Speaker could reconstruct the main event via visual cues.

(C) In the Reportative Access version, an Informant (e.g., the dog in panel C1) watched the entire event from the other side of the stage. The Informant saw the Agent stand next to the objects (e.g., the gorilla next to the board, panel C1), and manipulate the objects (coloring the star, panel C2 and C3). Later the Informant verbally informed the Speaker (e.g., the cat) about what had happened (panel C4; see the speech bubble with a snapshot of the event). During the act of reporting the stage curtains were closed so the Speaker would only have verbal information about the event.

All scenarios and versions followed the above basic structure. A complete list of the events in our stimuli can be found in Appendix A. For each of the roles of Agent, Speaker and Informant we used between 5 and 7 different animal puppets.

There was also a description of each event which appeared in a speech bubble next to the Speaker. The speech bubble appeared after the pictures within a scenario were displayed for 8 seconds and remained onscreen together with the pictures for 10 seconds. The event description was constructed in an artificial language (see Figure 1). As mentioned already, this language shared the English lexicon (but lacked determiners) and had a Subject-Object-Verb word order. We created two versions of each description: one with a novel verb-final morpheme, *ga*, that served as an evidentiality morpheme (*Gorilla star coloredga*) and one without it (*Gorilla star colored*).

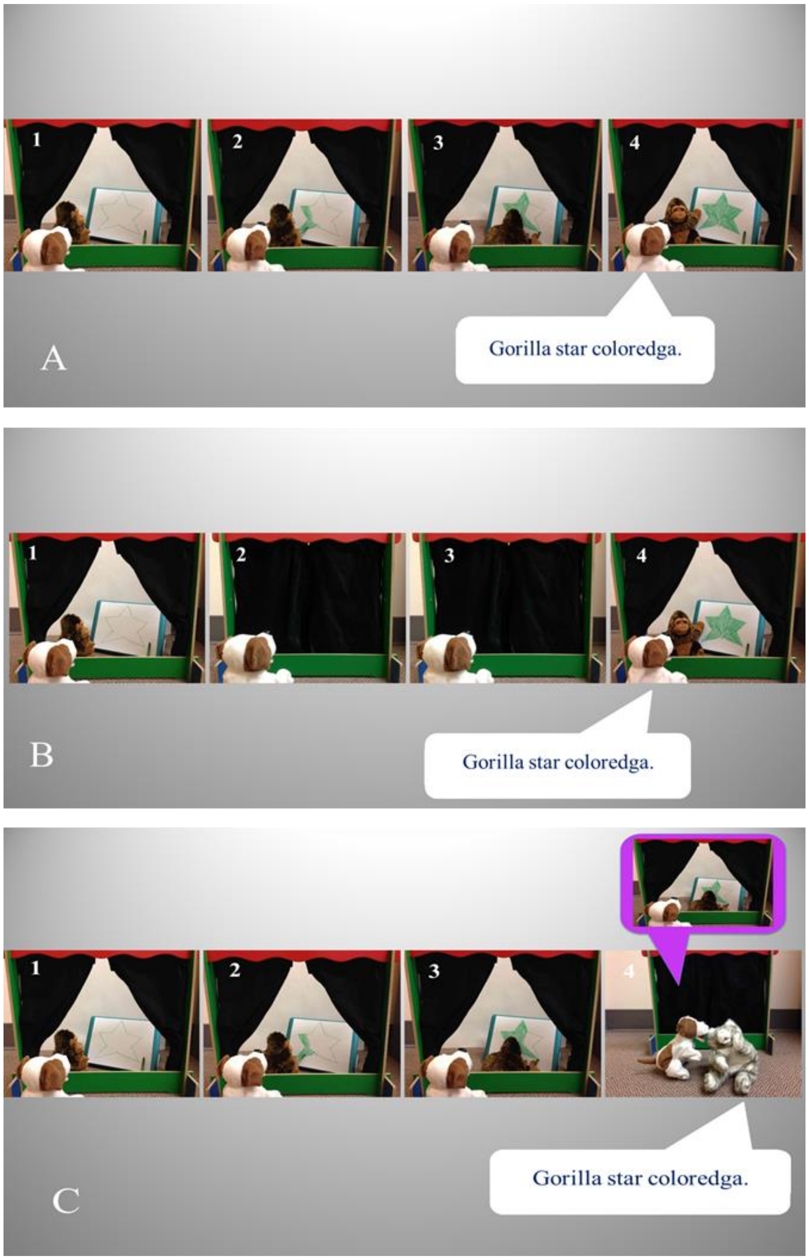

Figure 1. Sample stimuli for the Artificial Language Learning task of Experiment 1. Versions of the same scenario are shown in which a character gained access to an event in one of three ways: (A) Visual, (B) Inferential, and (C) Reportative. At the end, the main character offered a description of the event with or without an evidential (e.g., "Gorilla star coloredga.").

Of the 39 scenarios, 21 were used during the Training Phase and 18 during the Testing Phase of the experiment. Since each scenario had 3 versions depending on the

Speaker's Access to the event, 3 different lists were created for each Phase by selecting an equal number of scenarios from each Access type and rotating the scenarios across lists so that no list contained the same scenario in more than one version. For each list, there were 3 randomized presentation orders created, thus ending up with 9 presentation lists in total, each with 21 items for the Training Phase (7 from each Access type), and 18 for the Testing Phase (6 from each Access type). For the 18 scenarios in the Testing Phase, within each Access type, in half of the scenarios, the Speaker described the event erroneously used the morpheme *ga*, either failing to use the morpheme when it was needed or using it for the wrong type of Access. For the remaining half of the videos, the use of the morpheme was correct.

To instantiate different evidential systems during the Training Phase, only scenarios of a single Access type (Visual, Inferential, Reportative) per list were described with the morpheme *ga* to reflect one of three evidential systems (Visual, Inferential, or Reportative). Scenarios of the other two Access types were described by plain (unmarked) sentences during the Training Phase. Participants were randomly assigned to one of 3 between-subjects conditions depending on the evidential system they were exposed to during training (such that 37 participants were assigned to each evidentiality system).

2.1.1.3. Procedure

We tested participants in small groups in a dimly lit, quiet room. Stimuli were displayed using PowerPoint, Microsoft Office on a Windows laptop computer and projected on a large screen. In the beginning of the experiment, participants were given an individual response sheet and told that the experiment was comprised of two phases: a learning phase during which they would not have to write anything down but would only have to pay close attention to the pictures shown and try to figure out what was happening. In the second part, they would have to fill out their answers on the response forms following the instructions they would be given before the actual part began.

In the beginning of the Training Phase, participants were shown the following instructions:

> "In this experiment you will watch a series of events in a puppet theater. For each event, there will be a character outside the puppet theater that will learn about the event. Here is what the puppet theater will look like:"

At this point, a single picture of the puppet theatre was shown with two puppets, one on each side of the puppet theater in an effort to familiarize participants with the setting. Subsequently, participants were told that "each event would be depicted through four static clips shown next to each other from left to right in the order in which they unfold" and they were shown a sample event. This sample event was always the same and was structured in the same way as the main scenarios (a bear spilled candies while a donkey watched the complete event as the curtains of the puppet theater were open).

On the next instruction screen, participants received the following information:

> "At the end of each event, this outside character will offer a description of the event. The description will appear in a speech bubble next to the character. The language that the characters will be using is an alien language: it shares some words with English but is different in several ways. One difference is that the language includes a special word, "ga".

Subsequently, participants were told that they would have to pay attention "to when "ga" is and is not used in the language to figure out what it means" and that they would be given several events to go through in order to figure this out. Each instruction screen was displayed for approximately 20s. Then the display continued to the main part of the Training Phase of the experiment during which participants were exposed to events in which the morpheme *ga* was used to mark only a single type of Access (Visual, Inferential or Reportative) depending on the Evidentiality System that each group of participants was randomly assigned to. When the Training Phase was completed, participants were informed that they would continue to the second part of the experiment.

The Testing Phase began with the following instructions to participants:

> "We will now show you a new set of events in the same puppet theater. As before, for each event, an outside character will offer a description of the event in the alien language. But this time some of these descriptions will contain errors: they will contain "ga" when it is not appropriate or omit it when it is needed. For each event, please mark "Yes" or "No" on your answer sheet to indicate whether the character described the event using the alien language correctly."

Following the instruction screen, participants watched one scenario at a time and they were given 30 seconds at the end of each scenario to fill in their answer (the question of whether the character used *ga* correctly or not was repeated on this response form and participants had to indicate "Yes" or "No"). In the Testing Phase, half of the scenarios within each Access type were marked with *ga* and the rest were unmarked. This meant that half of the time the puppet used the alien language correctly (by including or omitting *ga* as appropriate) and the other half the puppet was incorrect by using the morpheme for the wrong type of Access or failing to use the morpheme for the target

Access type. When the experiment was over, participants were asked to write down what they believed *ga* meant and when it was or was not used.

2.1.2. Results

Participants' responses were coded for Accuracy, a binary outcome variable coded as 1 (Correct) or 0 (Incorrect). A summary of the data can be seen in Figure 2. The data were analyzed using logistic mixed-effects modeling (Baayen, 2008; Baayen, Davidson, & Bates, 2008). Since the data include categorical variables, the generalized binomial linear mixed effects modeling (glmer) function of the lme4 package was used for our analysis (Bates, Maechler, Bolker, & Walker, 2015) in the R Project for Statistical Computing (R Development Core Team, 2018).

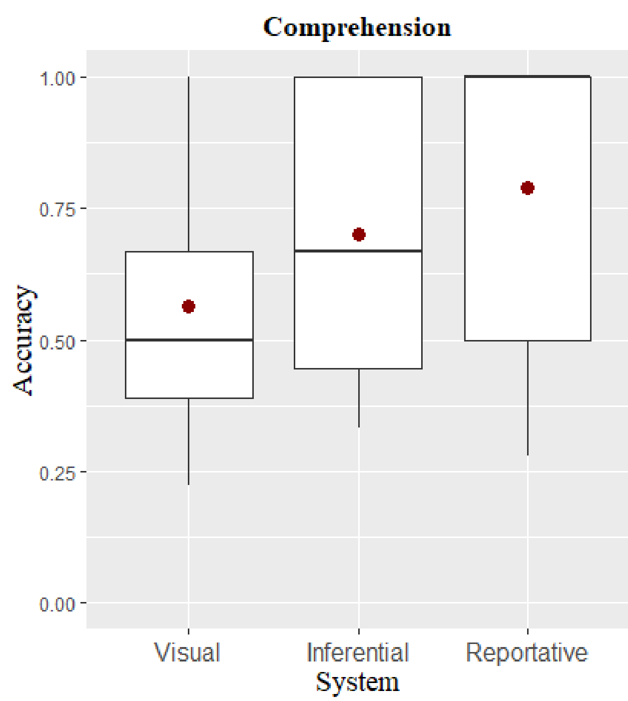

Figure 2. Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in the Artificial Language Learning task of Experiment 1.

Our mixed-effects model included Accuracy as the dependent variable and System (Visual, Inferential, Reportative) as our fixed predictor. To explore the possibility that non-theoretically driven predictors affected participants' performance, we included List (1, 2, 3) as a second fixed predictor along with its interaction with System. Our model also included random intercepts for Participants and Items. The fixed effect of System was assessed with two planned comparisons using contrast coding: the first contrast compared the Reportative system against the Visual and Inferential Systems and the second contrast compared the Visual and Inferential Systems to each other (contrast 1: 0.66, -0.33, -0.33, contrast 2: 0, 0.50, 0.50). The same contrast setup was used in all subsequent analyses. For the exploratory predictor List, we set up our contrasts as deviations from the overall mean, including no specific reference group since there was no theoretical justification to assume one (see Table 2).

The inclusion of System significantly improved the model fit based on a chi-square test of the change in -2 restricted log likelihood ($\chi^2 = 18.05$, $p = .0001$) but no such improvement was found for List ($\chi^2 = 3.10$, $p = .21$). Participants performed better in the Reportative System compared to the Visual and the Inferential Systems (M_{Rep}, = 0.74, $M_{Vis/Inf} = 0.63$). There was also a significant difference in performance between the Visual and Inferential Systems, with higher accuracy for the Inferential System ($M_{Vis} = 0.56$, $M_{Inf} = 0.70$). The inclusion of the interaction between System and List also improved model fit ($\chi^2 = 9.79$, $p = .04$) but the source of this interaction could not be determined by follow-up comparisons (see Table 2). When we reran the analysis reordering the levels for List (3, 2, 1), results did not change but there was an interaction such that, in List 3, accuracy in the Reportative System was not better than in the Visual/Inferential Systems (β = -1.6776, SE = 0.6267, z =-2.677, p<0.01).

Table 2. Parameter estimates for Accuracy in the Comprehension task of Experiment 1. Significance levels: ** p < .01, *** p < .001

Effects	Estimate	SE	z	
Intercept	1.64	0.2615	6.30	***
System (Reportative vs. Visual/Inferential)	1.80	0.47	3.81	***
System (Visual vs. Inferential)	-1.15	0.48	-2.37	***
List (List 1 vs. Mean)	0.14	0.29	0.50	
List (List 2 vs. Mean)	0.38	0.29	1.29	
System (Reportative vs. Visual/Inferential): List 1	0.80	0.66	1.22	
System (Visual vs. Inferential): List 1	1.08	0.67	1.59	
System (Reportative vs. Visual/Inferential): List 2	0.86	0.67	1.28	
System (Visual vs. Inferential): List 2	-0.37	0.68	-0.54	

Participants' answers to the post-experiment question about the meaning and use of *ga* confirmed this picture. Of the 37 participants exposed to the Visual System, only 5 correctly associated *ga* with complete visual access to the event and/or the verb *see*. Two participants incorrectly mentioned a different type of access. The majority (n = 30) associated *ga* with an irrelevant, i.e., non-evidential meaning (event completion, or other morphemes such as *the*, *-ed* or the plural *-s*). Of the 37 participants exposed to the Inferential System, 9 correctly mentioned that the speaker only saw the beginning and the end of the action; an additional participant mentioned doubt. The majority of the participants (n = 27) associated *ga* with irrelevant function morphemes (*the*, *-ed*, *it*, *them*).

Lastly, of the 37 participants exposed to the Reportative System, 17 participants correctly mentioned that the morpheme meant telling the other puppet what happened, and an additional participant linked its meaning to a secret. Four additional participants associated the morpheme with the act of one puppet talking to another (these answers were appropriate but did not unambiguously point to a mental-state meaning). The remaining 11 participants either associated *ga* with an irrelevant grammatical morpheme (*the, -ing, -ed*) or did not provide an answer. In sum, participants' overt conjectures about *ga* often revealed sensitivity to evidential interpretations (as opposed to other possible mental-state meanings such as uncertainty, or possibility, or other non-mental state interpretations). Furthermore, the distribution of correct evidential interpretations confirmed the advantage of the Reportative system. Finally, error patterns for incorrect evidential interpretations were consistent with the hypothesized aversion to mark visual access with a dedicated morpheme: the only cases where people adopted an evidential interpretation but chose the wrong source (n = 2) were observed for the Visual System.

2.1.3. Discussion

According to the TPH (Gentner & Bowerman, 2009), the most typologically common linguistic distinctions should be the easiest to learn. In an Artificial Language Learning experiment, we set out to explore whether this prediction is borne out within the semantic domain of evidentiality. Our results are in line with the TPH prediction: the most typologically common evidentiality system - the Reportative system - was learned more easily by our participants compared to the less common Visual system. The Inferential system was also learned better than the Visual system. This pattern was confirmed in participants' post-experiment responses: participants exposed to the Reportative System consistently associated the morpheme with the speaker's information access when explicitly reporting what the morpheme meant.

Even though the present experiment offers initial support for the TPH, it has a number of limitations. First, the use of static displays to depict events did not properly capture the way people access events in everyday life, since the sequence of pictures was composed of multiple 'snapshots' and always afforded a somewhat indirect path to the event. Second, and relatedly, properties of the static displays themselves might have influenced the results. For instance, instances of Reportative access were marked by a salient bubble containing a depiction of the event (see Figure 1) and were the only scenarios that contained three (as opposed to two) characters. These stimulus features might have made Reportative access more salient and might have facilitated participants' learning of the Reportative system. Third, Experiment 1 only included a comprehension test. It is possible that production of evidentials behaves differently from comprehension (see Ünal & Papafragou, 2016, among others). In Experiment 2 we addressed these limitations.

2.2. Experiment 2

Experiment 2 had the same logic and learnability targets (i.e., evidential systems) as Experiment 1 but used videos of unfolding events in which a character gained access to events in different ways instead of static scenarios. These videos were closer to the situations in which people find out about events in the world and were designed to offer a clearer separation of various information sources. Moreover, the visual characteristics of the videos were consistent across systems (e.g., all versions of the same video contained three people). This step was meant to ensure that any advantage found for one evidential system over the other could not be attributed to properties of the stimuli. Lastly, unlike Experiment 1, Experiment 2 tested both production and comprehension of the novel evidential morphemes.

2.2.1. Methods

2.2.1.1. Participants

We recruited 101 adult participants between the ages of 18 and 22 (73 female and 28 male). Participants were undergraduate students at the University of Delaware, enrolled in an Introductory Psychology course and earned course credit for their participation. None of them had participated in Experiment 1. All of them were native speakers of English. To the extent that participants were familiar with additional languages (mostly Spanish and French), these languages did not include grammaticalized evidentials.

2.2.1.2. Stimuli and Procedure

The experiment had a Training Phase and a Testing Phase (itself consisting of two tasks, a Production task and a Comprehension task). We filmed 69 videos in three versions each, with each version corresponding to a type of information access (Visual, Inferential, Reportative). Of these, 15 had some overlap with events from Experiment 1 and the rest were completely new. A complete list of the events in the videos for Experiment 2 can be found in Appendix B.

Each video displayed a different event with three characters. The roles of these characters were performed by the same female undergraduate research assistants and remained consistent across the videos. One of the characters performed an action using some materials and then put these materials away (the “Agent”). A second character gained access to the event in one of three ways (Visual Perception, Inference, Report), and at the end of the event, described what happened (the “Speaker”). The third character manipulated the Speaker’s access to the event by either allowing her to watch the event or by blocking her visual access for the complete duration or part of the event. Moreover, the setting in which the event took place was identical for all the videos: the Agent and the Speaker were sitting at opposite sides of a table while the third character stood behind the Speaker and had full view of the table. Each video was approximately 12 seconds long. At the end, the Speaker turned to the camera so as to describe what happened; at that point, the video stopped and the speech bubble with the artificial language sentence appeared, and stayed visible for 8 seconds before the next video began. The artificial language sentences were similar to those in Experiment 1.

Consider the sample event in Figure 3, in which the Agent copied a drawing on the whiteboard (all events in videos had a similar structure). In the Visual Access version in (A), the Speaker had continuous direct visual access to the event (we did include a brief initial phase in which the Speaker’s eyes were blocked, see panel A1, for visual consistency with the other versions). In the Inferential Access version in (B), the Speaker had visual access only for the beginning and aftermath of the event (panels B1 and B4), but not the main phase (panels B2 and B3); therefore, she had to infer what happened based on visual clues. In the Reportative Access version in (C), the Speaker's visual access was blocked while the event was unfolding (panels C1-C3); later (panel C4), the Speaker heard the third character's report about what happened. At the end of the video (identical across versions, see panel 5), the Speaker's description of what happened appeared in a speech bubble.

Figure 3. Sample stimuli for the Artificial Language Learning task in Experiment 2. We include sample screenshots from versions of a single dynamic video in which a character gained access to an event in one of three ways: (A) Visual, (B) Inferential, (C) Reportative. The three versions of the video unfolded somewhat differently (Panels 1-4) but had the same ending (Panel 5, enlarged here for readability). At that point, the main character offered a description of the event with or without an evidential (e.g., "She drawing copiedga.").

As in Experiment 1, we designed 3 evidential systems to be acquired (Visual, Inferential, Reportative) by having the Speaker describe only one type of Access with an evidentially marked sentence and include no evidential morpheme for the other two Access types. Participants were randomly assigned to one of these three evidential Systems (33 participants were assigned to the Visual System and 34 to each of the other two). We tested participants in small groups using the same set-up and procedure as in Experiment 1.

For the Training Phase, for each evidential system, we created three basic lists. Each list had 21 videos, 7 for each Access type. Across lists, each video rotated through each Access type. For each of these basic lists, we created three randomized presentation orders, resulting in nine lists in total. Each participant was exposed to one of these nine lists. The instructions participants received in the beginning of the Training Phase were similar to those in Experiment 1:

> "You will watch a series of short clips involving three characters: in each clip, one character will perform an action on some object(s) and then put the objects away. A second character will learn about this action in different ways depending on what a third character does. After this, the second character will describe what happened. This will be depicted in a speech bubble onscreen."

Participants were then shown a still image of the three characters in the visual scene shown in the videos so as to further familiarize them with the content of the stimuli before the Training Phase would begin. In the next instructions screen, participants were told:

> "The characters will be speaking an alien language: it shares some words with English but it is different in several ways. One difference is that the language includes a special morpheme: "ga". You will have to pay attention to when "ga" appears in order to try and figure out what it means. You will go through several events to try to learn the language and figure this out."

Before the videos started, participants were also informed that a second part would follow that would test their understanding of where *ga* appeared based on its meaning.

The Testing Phase began with a transition screen ("Let's continue") and proceeded to the Production and later the Comprehension Task. For the Production Task, we used 12 new videos, each filmed in each of the three Access types. We created three basic lists, each containing 12 videos, 4 per Access Type. As in the Training Phase, the Access version shown for each video was rotated across these three lists. For each basic list, three randomized presentation orders were created, resulting in 9 presentation lists in total. Each participant was exposed to one of the nine lists.

The structure of the scenarios displayed was identical to those shown in the Training Phase with the difference that when the speech bubble appeared at the end, the evidential morpheme was missing. At its place, attached to the verb, there was a gap and participants had to decide whether the morpheme should be used or whether the gap should remain empty. The specific instructions participants received were as follows:

> "We will now show you a new series of clips. Here you will be asked to help describe some events using the morpheme "ga" in the alien language. Almost everything will be the same as before, but this time no morpheme is included in the speech bubble. For each clip, in the blank Response box on your Response Sheet 1, please write in the verb with the morpheme that is needed or write only the verb if no morpheme is needed in order to complete the speech bubble."

Following these instructions, participants were given an example of how their answer should be using the sentence "She piano played_": they were told that they should either write “played” or “playedga” depending on whether they thought the morpheme was needed based on what they figured out from the first part of the experiment.

For the Comprehension task we created 36 new videos, each filmed in three versions to correspond to the three Access Types. As before, three basic lists were created by rotating the Access Type shown by each video. Subsequently, three unique randomized presentation orders were created for each list (9 presenation lists in total). Each list contained 36 scenarios, 18 per each Access type. In half of these 36 scenarios, the Speaker erroneously used the morpheme *ga*: she either failed to use the morpheme when she should or used it for the incorrect type of access. In the remaining half of the scenarios, the use of the morpheme was correct. Participants’ task would be to write “yes” or “no” in their response sheet depending on whether they thought the character was correctly used the morpheme or not. The instructions they received were the following:

> “We will now show you a third series of clips. Almost everything will be the same as when you paid attention to when “ga” is used, but this time some of character’s descriptions will contain errors: they will omit ‘ga” or include it when it is not needed. For each clip, please mark “Yes” or “No” on your Response Sheet 2 to indicate whether the character’s description included or omitted “ga” correctly.”

Similarly to Experiment 1, at the end of the session we asked participants to write down what they thought *ga* meant and when it was used or not used.

2.2.2. Results

Similarly to Experiment 1, participants' responses were coded for Accuracy for both Production and Comprehension. The results can be seen in Figure 4.

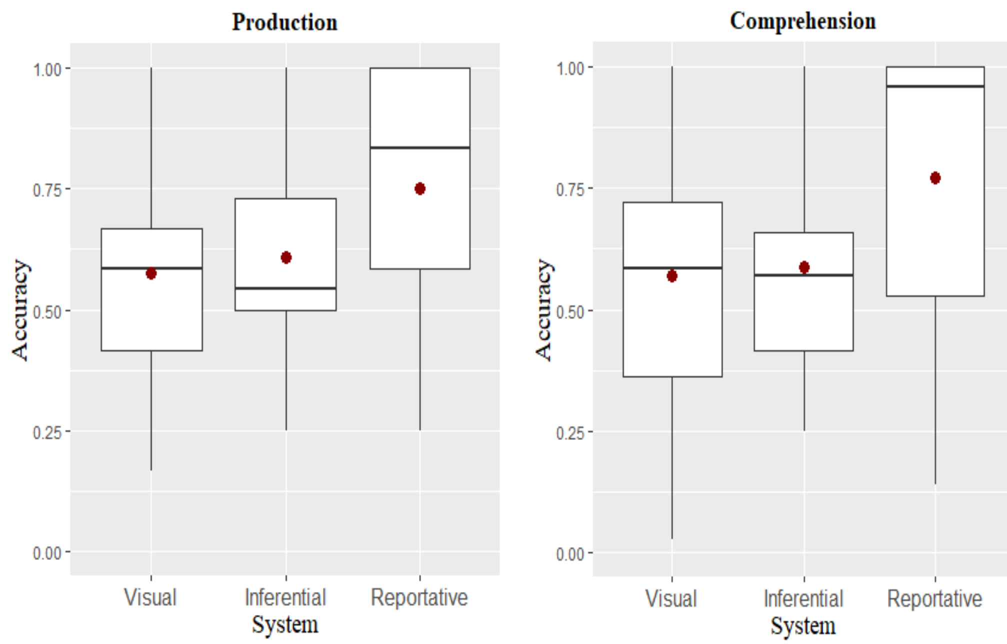

Figure 4. Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in the Artificial Language Learning tasks of Experiment 2.

For each of the tasks, we ran a mixed-effects model, using Accuracy as our dependent variable and System (Visual, Inferential, Reportative), List and their interaction as our fixed predictors. Our model also included random intercepts for Participants and Items. As in our previous analysis, we set up two planned comparisons using contrast coding such that we could compare the accuracy scores of the Reportative system to those of the other two systems, as well as the accuracy of the Visual and Inferential System (contrast 1: 0.66, -0.33, -0.33, contrast 2: 0, 0.50, 0.50, respectively; see Table 3).

For Production, our analysis revealed that including the fixed predictor System significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood (χ^2 = 10.40, *p*= .005). Participants' performance in the Reportative system was significantly better than their performance in the other two Systems (M_{Rep} = 0.75, $M_{Vis/Inf}$ = 0.59); there was no difference in performance between the Visual and Inferential systems (M_{Vis} = 0.57, M_{Inf} = 0.61). Model fit was not improved by the inclusion of either List (χ^2 = 3.67, *p*= .15) or the interaction between System and List (χ^2 = 3.96, *p*= .41).

For Comprehension, including System in our model significantly improved model fit (χ^2 = 16.16, *p* = 0.0003), but no improvement of the model was observed when List (χ^2 = 5.33, *p*= .07) or its interaction with System (χ^2 = 6.53, *p*= .16) were included. Following up on the effect of System, we found a significant difference between Reportative and the other two Systems (M_{Rep} = 0.77, $M_{Vis/Inf}$ = 0.58) but no significant difference between the Visual and Inferential Systems (M_{Vis} = 0.57, M_{Inf} = 0.59).

Table 3. Parameter estimates for Accuracy in the Production and Comprehension task of Experiment 2. Significance levels: ** p < .01, *** p < .001

Effects	Estimate	SE	z
<u>Production Task</u>			
Intercept	0.90	0.16	5.53 ***
System (Reportative vs. Visual/Inferential)	-1.10	0.34	-3.28 **
System (Visual vs. Inferential)	-0.16	0.37	-0.44
<u>Comprehension Task</u>			
Intercept	1.15	0.19	5.83 ***
System (Reportative vs. Visual/Inferential)	-1.73	0.42	-4.09 ***
System (Visual vs. Inferential)	-0.23	0.46	-0.49

Participants' answers when asked about the meaning/use of *ga* mirror this image. Of the 33 participants exposed to the Visual System, 12 correctly mentioned that *ga* was used when the speaker saw the event. Eight participants incorrectly associated *ga* with an alternative type of access (they mentioned that the girl's eyes were covered, the girl was told what happened, or something similar). The remaining 13 participants gave incorrect non-evidential conjectures (singular/plural forms, past or completed actions, articles such as *the*/*a*). Of the 34 learners of the Inferential System, 9 correctly associated *ga* with the character not seeing the complete action, or inferring the action, and an additional 2 mentioned doubt. Two participants had evidential but incorrect (reportative) meanings. The remaining 21 participants associated the morpheme with irrelevant grammatical morphemes (articles, *them*, singular/plural).

Lastly, of the 34 participants exposed to the Reportative System, 21 correctly associated *ga* with reportative access (importantly, these participants did not simply track the whispering or telling act from the perspective of the person committing the act but specifically alluded to the *speaker's* mental state by mentioning that she "was told" about the event). Three participants associated *ga* with the wrong source. The remaining 10 participants associated the morpheme with an irrelevant grammatical distinction. As in Experiment 1, these responses taken together confirmed that participants often chose evidential (as opposed to other possible) meanings for *ga*, did so more consistently in the Reportative System, and were more likely to reverse source meanings if asked to encode direct (visual) evidential meanings morphologically.

2.2.3. Discussion

In Experiment 2, we tested the learnability of evidential systems using videos displaying real-life, dynamic events. We also tested both receptive and productive knowledge of novel evidential morphemes. Our results reaffirm the TPH predictions: overall, the Reportative System was learned better compared to the cross-linguistically less-attested Visual and Inferential Systems. This learnability advantage is also supported by the large number of participants in the Reportative System condition that accurately associated the morpheme with the speaker's information access in a post-experiment probe.

Not all aspects of our data are compatible with the predictions of the TPH, however. Specifically, even though exclusive encoding of visual evidentials is rare, and there is a broad preference to mark non-visual/indirect over visual/direct sources cross-linguistically (Aikhenvald, 2018), the Inferential and Visual systems were equally learnable in our data. This pattern is different from Experiment 1. A possible explanation for this outcome lies with the fact that our Inference videos contained strong visual clues to what happened, bringing this type of information access closer to a direct perceptual experience than to an indirect inference on the speaker's part. This property might have blurred the distinction between Visual and Inferential access, a distinction that is subtle and somewhat language-dependent to begin with (de Haan, 2001, 2013), and might have led to a drop in performance with the Inferential system (cf. Figures 2 and 4).

By contrast, in Experiment 1, the Inferential scenarios were more distinct from the Visual ones because of the visual gap in the middle of the event (when the puppet theater curtains were drawn as the event was unfolding; see Figure 1). If so, replacing Inferential scenarios in Experiment 2 with less direct cases of inference from visual cues (e.g., by removing the agent; see Rissman, Woodward, & Goldin-Meadow, 2018) should increase the distance between visual and inferential access and might lead to a learnability difference between the Visual and Inferential systems.

Overall, Experiments 1 and 2 taken together show that there is a learnability difference among evidential systems that mirrors cross-linguistic tendencies to mark indirect information sources (especially, reports) before direct ones. This pattern raises the question whether this asymmetry is an exclusively linguistic fact or could generalize to other kinds of communication. In the next experiment, we turn to this question.

2.3. Experiment 3

2.3.1. Methods

2.3.1.1. Participants

We recruited 98 participants in total (age range: 18-22 years of age, 39 male and 59 female) from the same population as our previous experiments. None of them had participated in our earlier studies. All participants were native speakers of English. Some participants reported some knowledge of an additional language, mostly Spanish, French or Italian (none of these languages have grammaticalized evidentials).

2.3.1.2. Stimuli and Procedure

The stimuli and procedure were the same as in Experiment 2, with two modifications: (a) The speech bubble containing the witness' description of the event was in English instead of an artificial language (e.g., *She copied the drawing*); (b) the linguistic morpheme "ga" was replaced by a symbol (a black filled circle) that appeared next to the bubble and at the same time as the bubble. Both of these modifications can be seen in the example in Figure 5. The instructions were also adjusted to reflect these changes, as described below.

Figure 5. Example stimulus for the Artificial Symbol Learning task in Experiment 4 (Reportative System).

For the Training Phase, similarly to Experiment 2, we created three basic lists across which the Access version of each scenario was rotated through all three Access types (Perception, Inference and Report). For each basic list, three randomized presentation orders were created, resulting in nine lists in total, each of which contained 21 videos, 7 per Access Type.

The instructions in the Training Phase were the following (changes from Experiment 2 are in bold font):

> "You will watch a series of short clips involving three characters: in each clip, one character will perform an action on some object(s) and then put the objects away.

> A second character will learn about this action in different ways depending on what a third character does. After this, the second character will describe what happened. This will be depicted in a speech bubble onscreen. **Some clips will display a symbol on the screen in a separate box next to the speech bubble. This symbol is related to what is happening**. You will have to pay attention to **when this symbol appears** in order to try and figure out what it means. You will go through several events to try and figure this out."

Turning to the Testing Phase, for each task (Production and Comprehension), we used the three basic lists we constructed for Experiment 2, across which the Access version of the scenarios was rotated. For each of these lists, we created three new randomized presentation orders, resulting in 9 new presentations lists for each Task.

For the Production Task, each participant saw 12 scenarios, 4 per Access type. The instructions were modified (changes from Experiment 2 are indicated in bold font), such that participants had to produce a circle when appropriate (either draw a circle or, if desired, write the word "circle"):

> "We will now show you a new series of clips. Here you will be asked to fill in when appropriate **the same symbol** you have already seen. Almost everything will be the same as before BUT this time **no symbol is included** next to the speech bubble. For each clip, in the blank Response box on your Response Sheet 1, **please write in any symbol that is needed to complete the screen** or **cross out the blank** if no symbol is needed."

For the Comprehension Task, each participant saw 36 scenarios. In half of those scenarios, the symbol marked the correct Access type and in the remaining half, the symbol either appeared with the wrong Access type or was missing from the correct Access type. Participants had to write in the response form 'Yes' or 'No', depending on whether the symbol was included or omitted correctly. The instructions that participants received were as follows (once again modifications from Experiment 2 are in bold font):

> "We will now show you a third series of clips. Almost everything will be the same as when you paid attention to when **the symbol** [circle inserted here] appeared. BUT **some of the screens will contain errors:** they will omit it or include it when it is not needed.
>
> For each clip, please mark "Yes" or "No" on your Response Sheet 2 to indicate whether **the screen included or omitted the symbol** correctly."

As in Experiments 1 and 2, at the end of the session we asked participants to write down what they thought the symbol meant and when it was or was not used.

2.3.2. Results

Participants' responses were again coded for accuracy. The results can be seen in Figure 5. As for previous experiments, a mixed-effects logistic regression model was used, with Accuracy as the dependent, binomial variable, and System and List as the fixed predictors along with their interaction. The model also included random intercepts for Participants and Items (see Table 4).

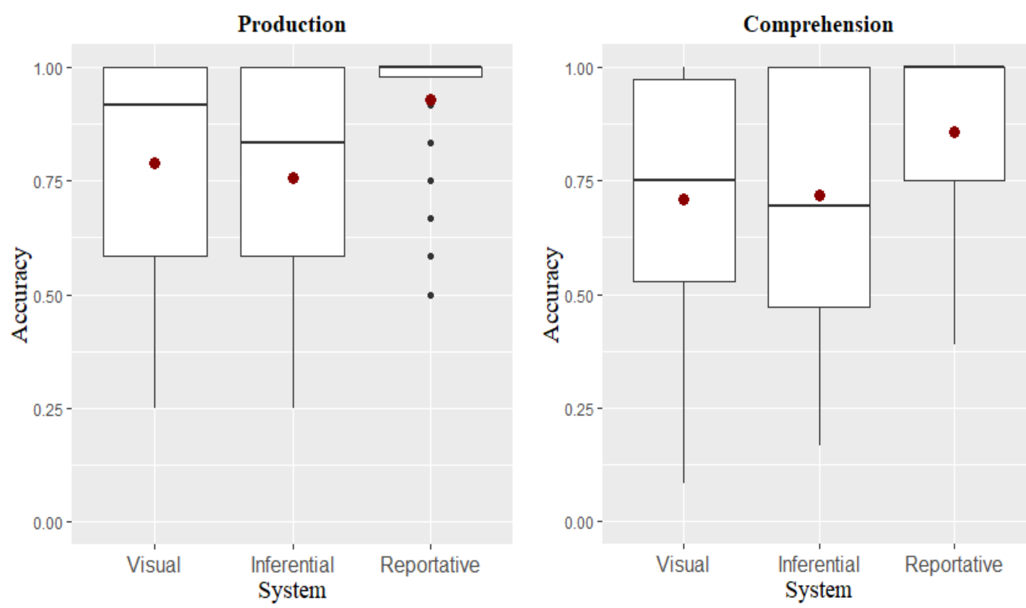

Figure 6. Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in the Artificial Symbol Learning tasks of Experiment 3.

Starting with the Production task, our analysis showed that only the fixed predictor System significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood (χ^2 = 14.99, *p*= .0005). Neither List (χ^2 = 3.92, *p*= .14) nor the interaction between System and List improved model fit (χ^2 = 5.34, *p*= .25). Following up on the role of System, there was a significant difference in performance between the Reportative and the other two Systems (M_{Rep} = 0.93, $M_{Vis/Inf}$ = 0.77) but there was no difference between the Visual and Inferential Systems (M_{Vis} = 0.79, M_{Inf} = 0.76).

For the Comprehension task, our model similarly showed that only System (χ^2 = 9.78, *p*= .007) but not List (χ^2 = 3.02, *p*= .22) or the interaction between System and List (χ^2 = 3.35, *p*= .50) improved model fit. Looking into the role of System, we found a significant difference between accuracy for Reportative when compared to the Visual and Inferential Systems (M_{Rep} = 0.86, $M_{Vis/Inf}$ = 0.71) and no difference in accuracy between the Visual and Inferential System (M_{Vis} = 0.71, M_{Inf} = 0.72).

Table 4. Parameter estimates for Accuracy in the Production and Comprehension task of Experiment 3. Significance levels: ** p < .01, *** p < .001

Effects	Estimate	SE	z
Production Task			
Intercept	2.94	0.35	8.25 ***
System (Reportative vs. Visual/Inferential)	-2.33	0.63	-3.66 ***
System (Visual vs. Inferential)	-0.39	0.61	0.63
Comprehension Task			
Intercept	5.57	0.31	8.15 ***
System (Reportative vs. Visual/Inferential)	-2.00	0.66	-3.03 ***
System (Visual vs. Inferential)	-0.17	0.66	-0.26

The debriefing data confirmed this picture. Of the 33 participants exposed to the Visual System, 17 gave correct conjectures (15 reported that the character saw the complete action and 2 said that character did not have her eyes covered). Three participants erroneously associated the symbol with reportative access. The remaining 11 participants gave incorrect answers (change in appearance of the objects, kind of motion). Of the 33 participants learning the Inferential System, 13 gave correct answers (11 mentioned that the character could only see the beginning and the end of the action and 2 that the character was blindfolded). Seven participants reported an incorrect evidential meaning (either saying that the symbol was used when the character could see "most or all the action" or associating the symbol with a report). The remaining 13 participants offered either incorrect answers (related with the type of action) or no answers. Finally, of the 32 participants exposed to the Reportative System, 21 correctly associated the symbol with the speaker getting informed about what happened by another person (all of them referred to the speaker's mental state and not only to the telling/whispering act). One additional participant associated the symbol with the character's lack of visual access. The remaining 10 participants gave irrelevant or no answers.

Lastly, we compared the results obtained from Experiments 2 and 3 that tested the learnability of evidential language and symbols respectively. Using a mixed-effect logistic regression model, as for the previous analyses, we included Accuracy as our dependent variable and System (Visual, Inferential, Reportative) and Experiment (Exp.2 - Language, Exp.3 - Symbol) as our fixed predictors. For this analysis, we excluded the predictor List since it did not significantly improve the model fit in any of the two

Experiments. We performed a separate analysis for each task across the two experiments (see Table 5).

For the Production task, both System (χ^2 = 21.37, *p*= .00002) and Experiment (χ^2 = 27.93, *p*= .0000001) but not their interaction (χ^2 = 2.07, *p*= .35) significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood. There was a significant difference in performance between the Reportative System and the other two Systems (M_{Rep} = 0.83, $M_{Vis/Inf}$ = 0.68) but no difference between the Visual and Inferential System (M_{Vis} = 0.68, M_{Inf} = 0.68). Participants were overall more accurate in Experiment 3 that involved Symbol learning compared to Experiment 2 that involved Language learning (M_{Exp1} = 0.64, M_{Exp2} = 0.82).

For the Comprehension task, we obtained similar results: System (χ^2 = 22.95, *p*= .00001) and Experiment (χ^2 = 11.75, *p*= .0006) significantly improved the model fit but their interaction did not (χ^2 = 0.02, *p*= .98). Follow-up analyses showed that participants performed better in the Reportative System compared to the other two Systems (M_{Rep} = 0.81, $M_{Vis/Inf}$ = 0.64) but no difference was found between the Visual and Inferential Systems (M_{Vis} = 0.64, M_{Inf} = 0.65). Participants were more accurate in Experiment 3 compared to Experiment 2 (M_{Exp2} = 0.64, M_{Exp3} = 0.76).

Table 5. Parameter estimates for Accuracy in the Production and Comprehension tasks of Experiments 2 and 3. Significance levels: ** p < .01, *** p < .001

Effects	Estimate	SE	z
Production Task			
Intercept	1.00	0.19	5.20 ***
System (Reportative vs. Visual/Inferential)	-1.52	0.31	-4.82 ***
System (Visual vs. Inferential)	-0.06	0.32	0.19
Experiment (Exp.2 vs. 3)	1.49	0.28	5.26 ***
Comprehension Task			
Intercept	1.22	0.22	5.35 ***
System (Reportative vs. Visual/Inferential)	-1.81	0.37	-4.91 ***
System (Visual vs. Inferential)	-0.19	0.39	-0.51
Experiment (Exp.2 vs. 3)	1.13	0.33	3.43 ***

2.3.3. Discussion

Performance on the Artificial Symbol Learning task of Experiment 3 was better compared to the corresponding Language Learning task of Experiment 2, presumably because of the costs of acquiring an unfamiliar language. Despite this difference, Experiment 3 replicated the basic finding from Experiment 2: the evidential system that marked only another speaker's report as source of information (Reportative) was learned better compared to both the evidential system that only marked inference (Inferential) and the system that only marked direct, perceptual experience (Visual). No significant difference was observed between the Inferential and the Visual systems.

We also see the same pattern in participants' reports about what the symbol meant: the vast majority of those exposed to the Reportative System associated the symbol with the speaker's epistemic state rather than the mere act of whispering or telling. We conclude that the bias to mark certain sources of information over others applies beyond language to other forms of ostensive communication. We discuss the nature of this bias and its effects on both typology and learnability in the following section.

2.4. General Discussion

2.4.1. Evidentiality and the learnability of semantic distinctions

A widespread assumption within cognitive science is that highly frequent cross-linguistic patterns should be easier to learn than less frequent ones because they reflect biases in human cognitive mechanisms. Within the domain of semantics, this idea has been captured by Gentner and Bowerman's (2009) TPH but has not received much empirical scrutiny. To our knowledge, our study was among the first (and very few) research studies that used an Artificial Language Learning methodology to explore learnability within the semantic domain (cf. also Xu, Dowman & Griffiths, 2013; Carstensen, Xu, Smith & Regier 2015; Maldonado & Culbertson, 2019). We compared the ease of acquisition of three evidential systems marking a single information source (Visual, Inferential, Reportative) by adult native speakers of English whose language lacks evidential morphology. Based on the relative frequency of these systems cross-linguistically, the TPH predicts that the Reportative should be the most learnable and the Visual system the least learnable of the three systems.

Overall, this prediction was partially supported. Across two different experiments with different stimuli (Experiments 1 and 2), the Reportative system was learned more easily compared to the Visual System. The learnability profile for the Inferential system was less consistent: in Experiment 1, the Inferential system was learned more easily than the Visual System but in Experiment 2, the Visual and the Inferential systems produced comparable learnability outcomes (and were both harder to learn than the Reportative system).

Together our data offer evidence for the conclusion that highly frequent semantic distinctions are more learnable than less frequent ones, as predicted by the TPH (Gentner & Bowerman, 2009). Furthermore, they add to previous studies that have studied learnability with the same methodological paradigm as ours mostly within the domains of syntax, phonology and morphology (Christiansen, 2000; Newport & Aslin, 2004; Hudson Kam & Newport, 2005, 2009; Seidl & Buckley, 2005; Wilson, 2006; Thompson & Newport, 2007; Wonnacott et al., 2008; Finley & Badecker, 2008; Tily et al., 2011; Merkx et al., 2011; Culbertson, 2012; Culbertson & Smolensky, 2012; Culbertson et al., 2012; Fedzechkina, et al., 2012; Tabullo et al., 2012).

2.4.2. Learnability and 'natural categories'

A striking aspect of our data was the finding that the preference to mark reportatives persisted when the participants' task was to acquire the evidential meaning of a pictorial symbol and not a linguistic morpheme (Experiment 3). This finding coheres with evidence from other non-linguistic communicative systems: many European and American communities that we are familiar with use a two-finger co-speech gesture for reportative attributions (e.g., *He wrote "the best essay in the whole school"*) but have no such gesture for visual or inferential attributions. Together, these results support the position that both the typological and the learnability patterns that characterize evidentiality are motivated by factors that go beyond language to characterize non-linguistic systems.

What could such factors be? Recall that, on Gentner and Bowerman's (2009) original proposal, the roots of the TPH lie in the cognitive naturalness of the semantic classes that the learner acquires: "All else being equal, within a given domain, the more frequently a given way of categorizing is found in the languages of the world, the more natural it is for human cognizers, hence the easier it will be for children to learn" (p. 467). One possibility, then, is that the present advantage of the Reportative system is the result of the naturalness of the underlying concepts.

Upon closer inspection of our data, this possibility seems surprising, even paradoxical. Within the class of evidential distinctions, it is visual perception – not reports - that could plausibly be considered as a primary, highly salient and valued source of information for humans and hence a primary candidate for both cross-linguistically consistent encoding and easy acquisition. For instance, perception as a knowledge source is represented in areas related to mental-state reasoning in both seeing and blind individuals (Koster-Hale, Bedny, Saxe, 2014). Some understanding that seeing leads to knowing emerges early in children (Pillow, 1989; Pratt & Bryant, 1990; Ozturk & Papafragou, 2016) and is even found in other primates (Hare, Call, & Tomasello, 2001). In memory, a common mistake is to think one has seen things with one's own eyes that were only indirectly experienced (e.g., imagined, inferred or read about; Johnson et al., 1993; Hannigan & Reinitz, 2001; Ünal et al., 2016); the opposite error rarely occurs.

Even within the domain of language, aside from evidential systems, the domain of vision is richly represented cross-linguistically with a variety of verbs (e.g., *see*, *perceive*, *look*; Viberg, 1984; San Roque et al., 2015; but see Majid et al., 2018), and the broad information-access meanings of these verbs seem to be accessible even in congenitally blind individuals (Landau & Gleitman, 1985). In sum, contrary to our findings, visual perception as an information source appears to be a 'natural' candidate for languages to mark and for learners to encode.

In what follows, we propose two distinct but not mutually exclusive possibilities to reconcile the prevalence of visual perception as a source of information for humans with the fact that such a prevalent source is neither easily acquired by learners acquiring a single evidential morpheme (or symbol) in our experiments nor preferentially marked in grammatical evidential systems cross-linguistically. Each of the two possibilities looks beyond the notion of naturalness as originally proposed in the context of the TPH. Together these two possibilities represent different, and probably complementary, principles that affect the nature and learnability of meaning distinctions within and beyond language.

2.4.3. Learnability and categories in the meaning hypothesis space

A first possible explanation for the learnability asymmetries we observed in both linguistic and non-linguistic (symbol) tasks attributes such asymmetries to the relative ease of splitting events into the right evidential categories (the 'category partition' explanation): events that were marked by the Reportative system were more clearly distinct as a class compared to events that were marked by the Visual and Inferential systems. This is because, as we have already mentioned in the discussion of Experiment 2, some instances of inference from perceptibles can be hard to distinguish from instances of perception proper. As a result, when learning the Visual or Inferential system, participants were required to learn to mark half of a set of events that were often indistinguishable from each other, which was likely to be difficult; when learning the Reportative system, all and only the events within a distinct class were marked, so learning to mark that category of events was easier.

The category partition explanation is in line with the observation in the linguistics literature that inferential evidentials have a mixed status, sometimes behaving like visuals and at other times behaving like reportatives (de Haan, 2001). It is also consistent with several findings from a recent study by Ünal, Pinto, Bunger and Papafragou (2016) suggesting that inference from visual experience is not a homogeneous category in cognition (see also Ünal & Papafragou, 2018). In one experiment of that study, English speakers were exposed to an event and had to describe it and state how they had found out about it. When participants had experienced an event directly (e.g., a woman drink milk), they stated having seen the event. However, when they had only seen visual cues from the aftermath of an event and had to "fill in" the rest, their statements varied. Closer inspection suggested that, when the visual cues were indeterminate (e.g., a woman next to a birthday cake, leading to the inference that she blew out the candles), participants consistently stated that they had inferred the event. By contrast, when the visual cues were more determinate and highly constrained the inference (e.g., a woman next to eggshells, leading to the inference that she cracked the egg), participants were equally likely to say that they had seen or inferred the event. The authors proposed that there are several varieties of inference, and that stronger, more constrained (and thus more secure) inferences from visual cues might be difficult to distinguish from purely perceptual experience.

In another experiment, Ünal et al. (2016) found that these varieties of inference had implications for the use of evidential morphology in a language that grammaticalizes evidentiality: speakers of Turkish used the indirect evidential in their language more consistently for the very same subset of events that English speakers had judged to be unequivocally inferred compared to those events that elicited more mixed source judgments. Furthermore, inference types had effects on memory: building on classic studies showing that people often have a false memory of having actually experienced events that they have only inferred (Johnson, Hashtroudi, & Lindsay, 1993; Hannigan & Reinitz, 2001; cf. Strickland & Keil, 2011), Ünal et al. (2016) found that, across English and Turkish speakers, such misattributions to perception were more common when inferences were strongly constrained by visual cues and thus harder to distinguish from pure perception.

The category partition explanation captures the typological prevalence of reportative systems, since such systems mark a category of source meanings that is easily separated off from the others (Plungian, 2001; Cornillie, Arrese & Wiemer, 2015). Furthermore, this explanation is reminiscent of the idea that observed patterns of semantic typology can be understood in terms of connectedness of categories in a discrete semantic map (Jurafsky, 1996; Haspelmath, 1997; van der Auwera & Plungian, 1998), or the relative positions of specific meaning distinctions within a continuous semantic space (Levinson & Meira, 2003; Majid, Boster & Bowerman, 2008; Gärdenfors, 2014).

2.4.4. Learnability and pragmatic marking

A second possibility is that the learnability biases (and the broad typological facts) in the semantic domain of evidentiality are related to the pragmatic implications carried by different sources of information, specifically the fact that direct or visual information sources are generally (even though not always) more reliable compared to non-visual or less direct sources (Dancy, 1985; Papafragou et al., 2007; Matsui & Fitneva, 2009; Aikhenvald, 2018; Wiemer, 2018), with reports being the least direct sources of all. We know that even young children track a speaker's trustworthiness and choose to learn things from reliable over unreliable speakers (Sabbagh & Baldwin, 2001; Koenig, & Harris, 2005; Jaswal, & Neely, 2006; Fusaro, Corriveau, & Harris, 2011; Harris, 2012; Koenig, 2012; Mascaro & Sperber, 2009; Jaswal, 2010). Both typology and learnability then tend to prioritize marking indirect, potentially less reliable information sources (especially verbal reports). This perspective coheres with proposals according to which human cognition is equipped with epistemic vigilance so to avoid unreliable sources and the risk of being misinformed (Sperber, Clement, Heintz, Mascaro, Mercier, Origgi & Wilson, 2010).

According to this 'pragmatic bias' hypothesis, non-perceptual sources in our data are selectively marked in systems that have a single evidential morpheme or symbol because they are *informative*, i.e., they represent a departure from the primacy of perception as an information source (Grice, 1989; Barnard, Rosen & Matthews, 2017). Furthermore, given that reportative access is the least direct type of information source (since the speaker need have experienced no part of an event), it is the most likely to be encoded in a system that has a single evidential marker. This perspective makes more substantive the conjecture that "the tendency to mark direct, or visual, or sensory evidentials less than others may reflect the primacy of vision as an information source" (Aikhenvald, 2018, p.16). From a theoretical viewpoint, this proposal is consistent with evidence that, early on in life, humans selectively comment on whatever they find unexpected (Greenfield & Smith, 1976; Greenfield, 1979), and even young children calculate informativeness as they learn (Frank & Goodman, 2014) and produce (Bannard et al., 2017) language. From an empirical viewpoint, this proposal is reminiscent of work showing that both young children and adults are more likely to interpret novel verbs in a story as mental verbs such as *think* when the verbs are used to describe *false* as opposed to true beliefs (Papafragou et al., 2007) – presumably because such beliefs have an unusual, and thus pragmatically notable, property (cf. also Hacquard & Lidz, 2019).

The pragmatic bias hypothesis predicts that visual evidentials should be learnable as long as they are part of an evidential system that also encodes some form of non-visual or indirect evidential (such two-way systems are encountered frequently in the world's languages; Aikhenvald, 2018). Existing developmental evidence supports this prediction (see Ünal & Papafragou, 2016; de Villiers et al., 2009). On this approach, then, what learners can and cannot learn in our data and beyond the lab is determined not only by what is easy or hard to conceptualize or categorize but also by what sorts of meaning distinctions are more or less informative and thus likely to be chosen for encoding by a communicator during an interactive exchange. This approach is compatible with other recent perspectives that have linked the organization of the lexicon cross-linguistically to communicative pressures related to efficiency (Gibson et al., 2017; Kemp, Xu & Regier, 2018; cf. also Haspelmath, 1997; Fedzechkina et al., 2012), and connects with pragmatically-inspired work in both adult psycholinguistics (Sperber, Cara & Girotto, 1995; Van der Henst, Bujakowska, Ciceron, & Noveck, 2006) and semantic development (Clark, 1990; Grigoroglou, Johanson & Papafragou, 2019).

2.4.5. Towards a theoretical synthesis

Each of the two hypotheses presented above goes beyond a simple mapping between 'natural' concepts and learnability outcomes as originally proposed by the TPH and instead attempts to explain learnability patterns in the domain of evidentiality in terms of which divisions of the meaning space are internally coherent to entertain (category partitions hypothesis) or informationally appropriate to convey (pragmatic bias hypothesis). Each of these hypotheses can explain cross-linguistically robust preferences to single out and encode specific classes of meaning over others, and can accommodate the fact that such preferences can also extend to meanings communicated by non-linguistic symbols. Furthermore, the two hypotheses are not mutually exclusive: for instance, even if the advantage of the Reportative system is explained pragmatically, one would need to invoke the category partitions hypothesis to explain why the Visual and Inferential systems behave similarly across our studies.

Up until this point, our findings leave open the precise contributions of these two hypotheses for the learnability of evidential distinctions but future experiments can throw light on their relative roles in a systematic way. For instance, the semantic partitions hypothesis predicts that learners should be equally successful with the current Reportative system and a mirror image of the system that uses a single morpheme to encode Inferential and Visual events and leaves Reportative events unmarked: the idea is that such a novel system would maintain the conceptual coherence of the classes of meaning to be learned. By contrast, the pragmatic bias hypothesis maintains that learning should asymmetrically favor the current Reportative system over its mirror image.

Chapter 3

PRAGMATIC EFFECTS ON THE LEARNABILITY OF EVIDENTIALS

Cross-linguistically, evidential systems that mark indirect (inference- or hearsay-based) access to information are more prevalent compared to systems that mark direct (perception-based) access (de Haan, 2013; Aikhenvald, 2003, 2004, 2018). Additionally, when languages only have a single evidential morpheme, this morpheme most often marks reportative information (or, less frequently, both inference and reports) while other types of information sources remain unmarked (Aikhenvald 2003, 2004, 2018). Systems that mark only direct (visual) access with dedicated grammatical devices but do not grammatically encode other source types are unattested (ibid.). These facts suggest a markedness hierarchy for evidentiality, with direct (visual) evidence being the unmarked case.

If (as suggested above) direct visual access to information is the least frequently marked morphological distinction in the domain of evidentiality, it should also be the least learnable. This expectation was borne out in our first series of Artificial Language Learning experiments in which participants were exposed to miniature morphological systems encoding a single evidential category (visual, inferential or reportative access to information). Results showed that the evidential system that marked reportative access was consistently easier for participants to learn and the system that marked visual access was the hardest to learn.

How can the learnability results be explained? On the one hand, these patterns sit uneasily with the idea that both typological frequency and learnability index the 'naturalness' of the underlying concepts (Gentner & Bowerman, 2009, a.o.), since visual perception is known to be a salient, important and natural conceptual category. Developmental evidence suggests that children from early on can reason about visual access and treat it as a reliable source of knowledge (Pillow, 1989; Pratt & Bryant, 1990; Ozturk & Papafragou, 2016). The link between seeing and knowing is also understood by non-human primates (Hare, Call, & Tomasello, 2001; Call & Carpenter, 2001) while it is also seems to be available to both seeing and blind individuals. Additionally, visual perception is lexically represented through verbs across different languages (e.g., Viberg, 1984; San Roque et al., 2015; 2018; Majid et al., 2018).

On an alternative explanation, both the typological facts about evidential systems and the corresponding learnability asymmetries can be related to the pragmatic implications carried by different sources of information. Direct visual access is generally (even though not always) considered to be more reliable due to the correspondence of the speaker's experience with reality compared to less direct sources that can depend on incomplete inferential premises or an informant's reliability (Dancy, 1985; Papafragou et al., 2007; Matsui & Fitneva, 2009; Koring & De Mulder, 2015; Aikhenvald, 2018; Wiemer, 2018). According to this pragmatic perspective, visual perception is a prevalent, reliable source of information for humans but when *one* evidential meaning is singled out to be encoded in a morpheme, it is unlikely to correspond to the predominant information source that humans rely on. Since evidential morphology always involves a first-person perspective (unlike lexical verbs that can be used to report other people's access in addition to own's own; Speas, 2018), humans mark the source of their own knowledge in an evidential device when it is *informative* to do so – when the source is newsworthy, that is, has novel cognitive consequences (e.g., it is unreliable). According to this perspective, non-perceptual sources are selectively marked in languages that have a single evidential morpheme because they are informative, i.e., they represent a departure from the primacy of perception as an information source (cf. Barnard, Rosen & Matthews, 2017, on a similar notion of informativeness). Furthermore, given that reportative access is the least direct type of information source (since the speaker need have experienced no part of an event), it would be the most likely to be encoded when languages have a single evidential morpheme. This perspective coheres with (and explains) the prior observation that "the tendency to mark direct, or visual, or sensory evidentials less than others may reflect the primacy of vision as an information source" (Aikhenvald, 2018, p.16).

This pragmatic explanation is consistent with proposals according to which human cognition is equipped with epistemic vigilance so as to avoid unreliable sources and the risk of being misinformed (Sperber, Clement, Heintz, Mascaro, Mercier, Origgi & Wilson, 2010). We know that even young children track a speaker's trustworthiness and choose to learn things from reliable over unreliable speakers (Sabbagh & Baldwin, 2001; Koenig, & Harris, 2005; Jaswal, & Neely, 2006; Mascaro & Sperber, 2009; Jaswal, 2010; Fusaro, Corriveau, & Harris, 2011; Harris, 2012; Koenig, 2012). However, constantly exercising epistemic vigilance could entail an additional processing cost: the assumption that human communication is presumed to be truthful and informative (Grice, 1989) would be violated and speakers would need to evaluate not only the actual information they receive but also their interlocutor's reliability and intentions. One way of providing essential information within a fully cooperative communicative context without overtaxing epistemic vigilance would be to only mark indirect, potentially unreliable – but not direct perceptual, presumably more reliable - sources of one's experience.

At present, the pragmatic explanation for the learnability of evidential distinctions remains untested. In Saratsli et. al. (2020), because of design complexities, the fact that the visual evidential was hard to learn did not uniquely point to pragmatic factors (but could have arisen, e.g., from the difficulty of differentiating visual from inferential access to information in the specific scenarios used in the study; cf. Johnson, Hashtroudi, & Lindsay, 1993; Hannigan & Reinitz, 2001; Ünal, Pinto, Bunger & Papafragou, 2016). In what follows, we put the pragmatic explanation for the acquisition of evidentiality to a direct test using a simple design. In Experiment 1, using an Artificial Language paradigm, we compared the learnability of only two evidential systems, one that marks reported information (Reportative System) and one that marks visual perception (Visual System). We reasoned that, even in this highly contrastive design, the pragmatic account predicts that the Reportative system should be easier to learn. In Experiment 2, we replicated the study but told participants explicitly that the meaning they had to acquire related to one's information access. According to the pragmatic account, this addition should not affect the learnability data: even if participants know the kind of meaning they should hypothesize, pragmatic reasons still disfavor visual evidentials. Finally, in Experiment 3, we asked whether the learnability patterns observed for evidential morphology would also generalize to non-linguistic (visual) markers of evidential categories. If pragmatic factors were responsible for the learnability of morphological distinctions, the results should persist as long as the task continues to involve a manifest communicative intention.

3.1. Experiment 4

3.1.1. Methods

3.1.1.1. Participants

We recruited 64 participants between the ages of 18 and 70. Almost all were undergraduate students at the University of Delaware that participated for course credit, except for 6 individuals recruited through Amazon Mechanical Turk after the subject pool closed. There was no difference in performance between the two samples. All participants were native speakers of English and none of them reported speaking an additional language at home or a language that included grammatical evidentials. Since evidentiality is not grammatically marked in English, having native English speakers as our participants in this and the following experiments ensured that there would be no native language interference on learnability patterns.

3.1.1.2. Stimuli

Our stimuli were comprised of 42 videos (a subset of the videos used in Saratsli et al., 2020). Each video had two versions, with each version corresponding to a type of information access (Visual or Reportative). A complete list of the events in the videos for this experiment can be found in Appendix C. Each video depicted an event involving three characters. One of the characters performed an action ("the Agent") while the second character either fully witnessed this action (Visual Access version) or was informed about what happened without having any type of visual experience of the event (Reportative Access version). At the end of the video, this second character ("the Speaker") turned to the camera and described what happened. The third character manipulated whether the Speaker has access to the event by either allowing her to watch what happened or by blocking her visual access throughout the event. The same three, female undergraduate research assistants appeared in all videos and their roles remained constant throughout the different events. The event setting was also kept identical across the different videos with the Agent and the Speaker sitting at the opposite sides of a table while the third character was standing behind the Speaker.

An example of an event can be seen in Figure 7. For the Visual Access (Panel section A), the Speaker had continuous direct visual access while the Agent performed the event (she lit a lamp, panels A2-A4). To make these videos comparable to the Reportative Access ones, the Speaker's eyes were blocked by the third character in the beginning of all the Visual Access events (A1) but then the Speaker gained Access to the complete event (A2-4). For the Reportative Access (Panel section B), the Speaker's eyes were blocked throughout the event by the third character (B1-3). After the Agent had completed the event and took the materials away from the table, the third character uncovered the Speaker's eyes and was shown whispering to her (B4). For both access types, the video ended with the Speaker turning to the camera to describe what happened (Panel 5). At that time, a speech bubble appeared, containing a sentence in the target artificial language and remained on the screen for approximately 8 seconds before the next video started.

The artificial language had Subject-Object-Verb word order and lacked determiners but otherwise contained English words. The sentences in that language either contained a novel evidential morpheme, *ga*, at the end of the verb, marking the character's access to the event (*She lamp litga*) or omitted the evidential (*She lamp lit*). We constructed two evidential systems (Visual and Reportative) depending on which type of Access was marked with *ga*; within each system, the events corresponding to the other type of access were described using plain (unmarked) sentences.

Figure 7. Sample screenshots from versions of a single dynamic video in Experiment 4 in which a character gained access to an event of someone lighting a lamp either through visual access (A) or verbal report (B).

Our experimental paradigm consisted of a Training Phase in which participants were exposed to the events and the use of the artificial language and a subsequent Testing Phase, in which participants' ability to use (Production Task) and interpret (Comprehension Task) the novel morpheme was assessed. As mentioned before, each event was filmed in two versions depending on the Speaker's Access to the event (Visual or Reportative). This allowed us to create two counterbalancing lists by rotating the Speaker's access to the event (i.e., if a certain video in list 1 showed the Speaker having Visual Access to the event, the **same** video in list 2 would show the Speaker having Reportative access to the same event). For each list, we created two different presentation orders for each evidential system, resulting in four different randomized presentation lists per evidential system. Participants were randomly assigned to one of the two evidential systems and to one of the four corresponding presentation lists of that evidential system.

3.1.1.3. Procedure

Participants were tested through the use of one of two online platforms. Participants that received course credit were invited to connect through a Zoom session. Stimuli were displayed on the researcher's computer and screen shared with participants. Prior to the experimental session, participants were provided with an individual response sheet that they needed to fill out on Qualtrics at the time of the study. For participants that were recruited through Amazon Mechanical Turk (https://www.mturk.com/), the procedure remained the same but individual sessions were administered through Ibex PennController (Zehr & Schwarz, 2018).

In the beginning of the experimental session, participants were provided with a detailed description of the setting of the events:

> “You will watch a series of short clips involving three characters: in each clip, one character will perform an action on some object(s) and then put the objects away. A second character will learn about this action in different ways depending on what a third character does. After this, the second character will describe what happened. This will be depicted in a speech bubble onscreen.”

Next, a still image of the three characters in the visual setting was displayed in an effort to familiarize participants with the content of the stimuli before the actual videos began. Participants were then provided with the following instructions for the Training Phase, including details about the artificial language to which they would be exposed and what they needed to do:

> “The characters will be speaking an alien language: it shares some words with English but it is different in several ways. One difference is that the language includes a special marker: "ga". You will have to pay attention to when "ga" appears in order to try and figure out what it means. You will go through several events to try to learn the language and figure this out.”

Before the videos started, participants were also informed that there would be a second part in which their understanding of where *ga* appeared based on its meaning would be assessed. The Training Phase included 10 videos in total, 5 videos for each Access type.

After this phase, participants were first presented with the Production and later the Comprehension Task. For the Production Task, we used 8 new videos, 4 videos for each type of Access (Visual or Reportative). The setting and structure of the videos displayed was identical to the videos shown in the Training Phase but when the speech bubble appeared at the end of the video, there was a gap next to the verb where the evidential marker could be placed and participants had to decide whether the marker should be used or omitted. The specific instructions that participants read were as follows:

> "We will now show you a new series of clips. Here you will be asked to help describe some events using the marker "ga" in the alien language. Almost everything will be the same as before, but this time no marker is included in the speech bubble.
>
> For each clip, there is a textbox. Please type in either the verb with the marker that is needed or only the verb if no marker is needed in order to complete the speech bubble."

After given some time to thoroughly read and fully understand the instructions (approximately 15 seconds), participants were shown a sentence example that did not appear in the actual task so as for them to get a more precise idea of what their answer should be. This model sentence was: "She piano played_" and participants were explicitly told that they should type either the plain verb ("played") or the verb along with the marker ("playedga") depending on whether they thought the marker was needed for each given trial considering what they figured out from the first part of the experiment.

For the Comprehension task, we used 24 videos, 12 per each Access type. For half of the events the Speaker erroneously used the morpheme *ga*, either by omitting the morpheme when it should have been used or by using it for the wrong type of access. For the remaining half of the scenarios, the use of the morpheme was correct. Participants had to type in either "yes" or "no" depending on whether they thought the character was correctly using the morpheme or not. The specific instructions participants were given were the following:

> "We will now show you a third series of clips. Almost everything will be the same as when you paid attention to when "ga" is used, but this time some of character's descriptions will contain errors: they will omit 'ga" or include it when it is not needed.
>
> For each clip, please type **<u>Yes</u>** <u>or</u> **<u>No</u>** in the textbox provided to indicate whether the character's description **correctly** included or omitted "ga"."

After completing the Comprehension Task, participants were transferred to a new page where they were asked to provide a brief description of what they thought the marker *ga* meant and when it was used.

3.1.2. Results

Participants' responses were coded as a binary outcome variable (using 1 for a correct response or 0 for an incorrect response), reflecting their accuracy in each task. The data summary for each task can be seen in Figure 8. A logistic mixed-effects modeling was used for the data (Baayen, 2008; Baayen, Davidson, & Bates, 2008), implementing the generalized binomial linear mixed effects modeling (glmer) function of the lme4 package due to the presence of categorical variables (Bates, Maechler, Bolker, & Walker, 2015). The data analysis was performed using the R Project for Statistical Computing (R Development Core Team, 2018).

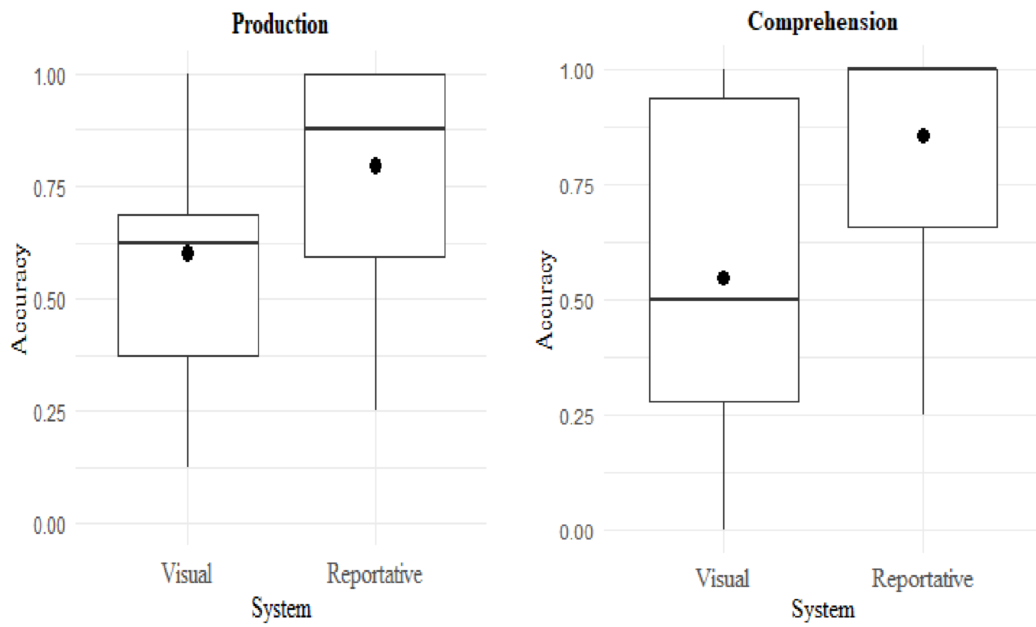

Figure 8. Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in the Production and Comprehension tasks of Experiment 4.

For each of the two tasks, the mixed-effects model included participants' accuracy as the dependent variable and the type of evidential System (Visual, Reportative) as a fixed predictor along with random intercepts for Participants and Items. Since the fixed predictor of System only had two levels, its effect was assessed with one planned comparison between the Reportative and the Visual systems (contrast coding: -.50,.50).

For the Production task, including the fixed predictor in the model significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood (χ^2 = 10.58, p= .001). Participants showed higher accuracy rates for the Reportative compared to the Visual System (M_{Rep}, = 0.79, M_{Vis} = 0.60). For the Comprehension task, we obtained similar results, with System significantly improving the model fit (χ^2 = 16.59, p< .001): the Reportative System yielded higher accuracy rates compared to the Visual System (M_{Rep}, = 0.85, M_{Vis} = 0.54). The parameter estimates of our models for both the Production and the Comprehension tasks can be found in Table 6.

Table 6. Parameter estimates for Accuracy in Experiment 4. Significance levels:*p<.05,** p < .01, *** p < .001

Effects	Estimate	SE	z
Production Task			
Intercept	1.18	0.21	5.41 ***
System (Reportative vs. Visual)	-1.26	0.39	-3.23 **
Comprehension Task			
Intercept	2.70	0.54	5.00 ***
System (Reportative vs. Visual)	-3.82	0.98	-3.87 ***

Participants' answers to the explicit question about the meaning and use of *ga* align with the accuracy rates observed in the two tasks. Ten of the 32 participants that were exposed to the Visual System correctly associated the morpheme with visual access to the event, specifically mentioning the verbs *see, watch* and *look*. Additionally, one participant did not directly associate the marker with visual evidence but instead with the fact that the third character did not whisper to the Speaker when the marker was used. Eight participants mistakenly associated the marker with the opposite (i.e., Reportative) access. The remaining 18 participants associated *ga* with irrelevant, non-evidential meanings (e.g., they mentioned other morphemes such as *the, a, -ed*, or other kinds of meanings, such as event completion). Of the 32 participants exposed to the Reportative System, 23 offered appropriate conjectures about the meaning of *ga*:15 participants associated the marker with the Speaker not knowing about the event and being told about what happened, 7 focused on the Speaker's lack of visual access, and the remaining 2 associated the marker with uncertainty and secret action. The rest of the participants provided meanings unrelated to evidentiality.

3.1.3. Discussion

It has long been assumed in the literature that, across languages, semantic distinctions that are more prevalent are also easier to learn. Our previous set of experiments in the have provided support for this assumption but have not determined the driving force behind the learnability patterns (Saratsli et al., 2020). Here we asked whether the ease of acquisition of evidential distinctions can be attributed to pragmatic (as opposed to conceptual) forces.

We exposed participants to one of two miniature evidential systems in a highly contrastive experimental setting with only two information sources. Our results show that the most typologically common (Reportative) evidential system was learned more easily compared to the less common (Visual) system. This pattern was confirmed by participants' explicit conjectures about what the marker meant: participants exposed to the Reportative System were more likely to explicitly associate the marker with the correct evidential meaning compared to participants exposed to the Visual System. Because reportative access is more informative compared to visual access, the current experiment offers support for the effect of pragmatic pressures on the learnability of semantic distinctions.

3.2. Experiment 5

If a pragmatic asymmetry within sub-types of evidential meanings drives the findings from Experiment 4, the learnability patterns should persist even if learners are explicitly told that the new morpheme encodes source of information. In Experiment 5, we replicated Experiment 4 but modified the instructions prior to the Training Phase to explicitly inform participants that the marker *ga* encoded the way that the speaker obtained information about an event.

3.2.1. Methods

3.2.1.1. Participants

We recruited 64 adult participants between the ages of 18 and 70. Participants were recruited mostly through the undergraduate subject pool at the University of Delaware and earned course credit for their participation (n = 52). A few were recruited through Amazon Mechanical Turk after the subject pool closed and received monetary compensation for their participation (n = 12). There was no difference in performance between the two sample groups. None of the participants had participated in Experiment 1. All were native speakers of English. Most of the participants did not speak a second language; in case participants were familiar with additional languages (mostly Spanish and French), these languages did not include grammatical evidentials.

3.2.1.2. Stimuli and Procedure

The stimuli and procedure were identical to Experiment 4. Participants were randomly assigned to one of two evidential systems (n = 30 for the Visual System and n = 29 for the Reportative System). We only modified the instructions participant received prior to the Training Phase in the following way (marked in bold font):

> "The characters will be speaking an alien language: it shares some words with English but it is different in several ways. One difference is that the language includes a special marker: "ga". **This marker shows how a speaker gets to know about what happened.**
>
> You will have to pay attention to when "ga" appears in order to try and figure out what it means. You will go through several events to try to learn the language and figure this out."

3.2.3. Results

Participants' responses were coded for Accuracy for both Production and Comprehension. The summary results can be seen in Figure 9.

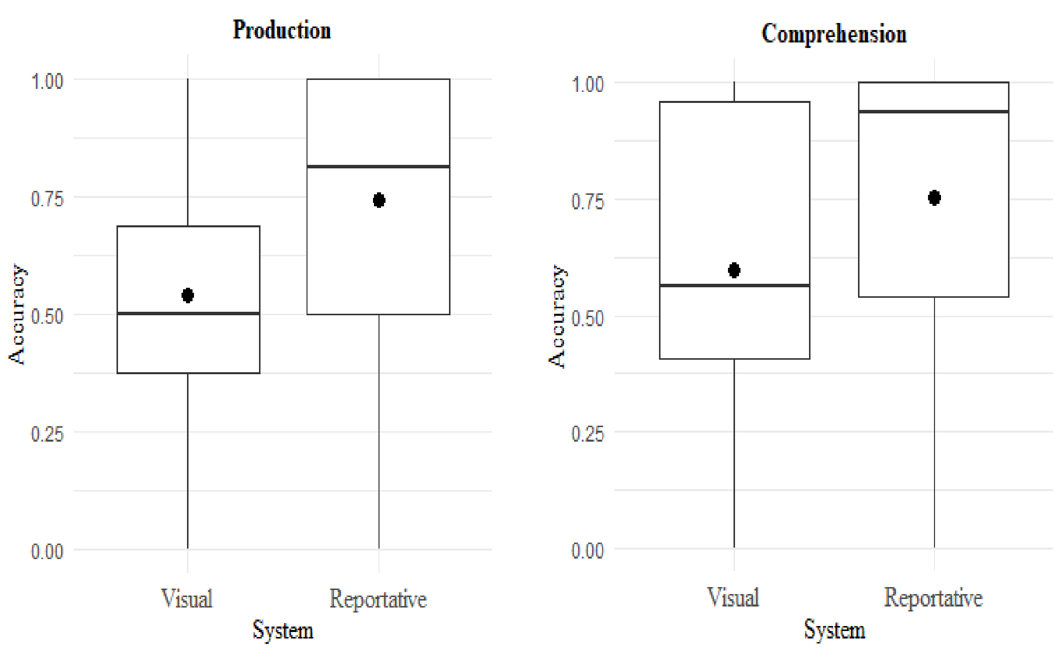

Figure 9. Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in Experiment 5.

A mixed-effects model was built for each of the two tasks, using participants' Accuracy as the dependent variable and the evidential System (Visual, Reportative) as the fixed predictor. Each model also included random intercepts for Participants and Items. We used contrast coding to compare the accuracy scores of the Reportative System to those of the Visual System (-.50,.50). Our analysis for the Production task revealed that the fixed predictor System significantly improved the model fit based on a chi-square test of the change in -2 restricted log likelihood (χ^2 = 7.02, *p*= .008): participants' accuracy rates in the Reportative System were higher than the accuracy rates observed in the Visual System (M_{Rep} = 0.74, M_{Vis} = 0.54). Turning to the Comprehension task, including System in our model improved model fit (χ^2 = 5.10, *p* = 0.02): again accuracy was higher in the Reportative compared to the Visual System (M_{Rep} = 0.75, M_{Vis} = 0.60). All parameter estimates for each model can be found in Table 7.

Table 7. Parameter estimates for Accuracy in Experiment 5. Significance levels: *p<.05,** p < .01, *** p < .001

Effects	Estimate	SE	z
Production Task			
Intercept	1.01	0.27	3.70 ***
System (Reportative vs. Visual)	-1.38	0.52	-2.64 *
Comprehension Task			
Intercept	1.68	0.39	4.26 ***
System (Reportative vs. Visual)	-1.69	0.75	-2.23 *

The participants' explicit conjectures about the marker's usage revealed a pattern similar to that of Experiment 4. Of the 32 participants exposed to the Visual System, only 11 associated the marker with the character's visual access to the event. Four participants incorrectly associated the marker with the opposite (i.e., Reportative) access, and the rest mentioned meanings unrelated to evidentiality (past or completed actions, articles such as *the*/*a*, and the plural form for the noun included in the sentence). Of the 32 participants exposed to the Reportative System, 14 correctly associated *ga* with reportative access, specifically mentioning that the marker was used when the speaker "was told" about the event, and another 4 mentioned the speaker's lack of visual access. The remaining participants associated the morpheme with irrelevant meanings.

Finally, we compared the accuracy rates for the Production and the Comprehension task separately across Experiments 4 and 5 (see Table 8). Starting with the Production task, we used a mixed-effect logistic regression model, with Accuracy as the dependent variable, System (Visual, Reportative) and Experiment (Exp.4 – Original Instructions, Exp.5 – Modified Instructions) as fixed predictors, and random intercepts for Items and Participants. Our results showed that System significantly improved our model based on a chi- square test of the change in -2 restricted log likelihood (χ^2 = 16.99, p< .001); Experiment (χ^2 = 1.18, p= .27) and the interaction between the two fixed factors (χ^2 =0.01, p= .90) led to no such improvement. This shows that participants' production was similar across the two experiments despite the modified instructions of Experiment 5 and only differed depending on the System they learned (M_{REP} = 0.76, M_{VIS}= 0.57).

Table 8. Parameter estimates for Accuracy in the Production and Comprehension tasks of Experiments 4 and 5. Significance levels: * p < .05, ** p < .01, *** p < .001

Effects	Estimate	SE	z	
Production Task				
Intercept	1.09	0.16	6.44	***
System (Reportative vs. Visual)	-1.32	0.32	-4.13	***
Experiment (1 vs. 2)	-0.34	0.31	-1.09	
System* Experiment	0.07	0.62	0.11	
Comprehension Task				
Intercept	2.12	0.31	6.69	***
System (Reportative vs. Visual)	-2.65	0.59	-4.45	***
Experiment (1 vs. 2)	-0.77	0.57	-1.34	
System* Experiment	1.82	1.15	1.58	

Turning to the Comprehension task, we used a similar mixed-effects regression model, including System and Experiment as fixed predictors and random intercepts for Items and Participants. As with the Production task, only System improved the model fit (χ^2 = 19.74, p< .0001) but Experiment (χ^2 = 1.43, p= .23) or its interaction with System (χ^2 = 2.50, p= .11) did not. Again, participants' performance was different across evidential Systems (M_{REP} = 0.80, M_{VIS}=0.57) but not across the two experiments.

3.2.4. Discussion

In Experiment 5, we used the same stimuli and procedure as in Experiment 1 but included an explicit clue about the evidential nature of the target morpheme. Our results replicated the learnability patterns obtained in Experiment 4: participants learned the Reportative System more easily compared to the (cross-linguistically less-attested) Visual System. Moreover, the clue included in the instructions did not boost participants' performance compared to Experiment 4. These findings strengthen the conclusion that the learnability advantage of Reportative over Visual evidential systems has a pragmatic explanation: even when the difficulty of identifying the abstract category of evidentiality as a candidate for learning was removed, participants still learned to encode certain kinds of information sources more easily compared to others – presumably because these meanings and their associated communicative contribution provided a more appropriate basis for linguistic marking.

3.3. Experiment 6

If the advantage of the Reportative System in our findings so far is due to pragmatic factors, the same advantage should occur even if the evidential distinctions were encoded through non-linguistic stimuli. This is because pragmatic principles are expected to characterize both linguistic and non-linguistic (e.g., pictorial) communicative systems (Grice, 1975; Sperber & Wilson, 1986; cf. also Richards, Kampa & Papafragou, 2019).

In Experiment 3, we addressed this possibility. This experiment adopted the stimuli and structure of Experiment 1 but included no artificial language. Instead of a sentence with an evidential morpheme, a red frame appeared around the video (and instead of an unmarked sentence, a plain video without a red frame was shown). The participants' task was to identify the type of event that was marked by a red frame. The speaker still described the event, but this description was in English with no evidential marker (e.g., *She lit the lamp*). Therefore, this linguistic content contained no information that could assist participants in tracking which events get marked. If pragmatic factors were responsible for the previous learnability data in Experiments 1 and 2, the same factors should still lead to a Reportative category advantage since the task still involved a manifest communicative intention (albeit one that was expressed through visual means).

3.3.1. Methods

3.3.1.1. Participants

We recruited 59 participants in total between the ages of 18 and 70. None of them had participated in our earlier studies. All participants were native speakers of English and were recruited through Amazon Mechanical Turk. None of the participants reported knowing a language that has grammatical evidentials.

3.3.1.2. Stimuli and Procedure

The stimuli and procedure were the same as in Experiment 1 but the speech bubble contained the description of the event in English with no evidential marker. In the Training Phase, for videos where *ga* would have been uttered, a red frame appeared around the events for the complete duration of the video. The instructions that participants received in the Training Phase were the following (changes from Experiment 1 are marked in bold font):

> "You will watch a series of short clips involving three characters: in each clip, one character will perform an action on some object(s) and then put the objects away. A second character will learn about this action in different ways depending on what a third character does. **Some of the events will have a red frame. You will have to pay attention to when the red frame appears in order to try and figure out what kind of event gets this frame**. **After this, we will show you two new series of events to see whether you figured out what kind of event gets a red frame**."

For the Production Task, participants had to indicate whether a frame should accompany each event within a new set of videos. The instructions were as follows:

> "We will now show you a new series of clips. Here you will be asked to indicate which events get a frame and which ones do not. Almost everything will be the same as before BUT this time no frame is included. For each event, please type "Frame" if you think that the event should have a red frame or "No frame" if you think that no frame is needed."

In the Comprehension Task, participants watched a set of videos, in half of which the frame appeared for the correct Access type while in the remaining half, the frame either marked the wrong Access type or was missing for the relevant Access type. Participants were asked to type in 'Yes' or 'No', depending on whether the frame was included or omitted correctly. The instructions were as follows:

> "We will now show you a third series of clips. Almost everything will be the same as before. BUT this time, some events will have a red frame when it is not appropriate, or some events will be missing the red frame when it is needed.
>
> For each event, please type "Yes" or "No" in the textbox provided for each event to indicate whether the presence or absence of the red frame is correct."

At the end of the session, participants were asked when the red frame was used and what it could mean.

3.3.2. Results

Participants' responses were coded for accuracy. The summary results for each task can be seen in Figure 10. Separately for each task, a mixed-effects logistic regression model was used, including Accuracy as the dependent, binomial variable, and Evidential System as the fixed predictor along with random intercepts for Participants and Items (Table 9). Starting with the Production task, the fixed predictor System did not significantly improve the model fit based on a chi- square test of the change in -2 restricted log likelihood (χ^2 = 1.90, *p*= .16). Even though numerically there was a learnability trend similar to the pattern observed for Experiments 1 and 2, with participants' accuracy for the Reportative System being higher than that of the Visual System (M_{Rep} = 0.89, M_{Vis} = 0.66), this difference did not reach significance - possibly because of the variance observed in participants' responses. Turning to the Comprehension task, the factor System did improve model fit (χ^2 = 7.54, *p*= .006), with participants showing higher accuracy for the Reportative than the Visual System (M_{Rep} = 0.90, M_{Vis} = 0.68).

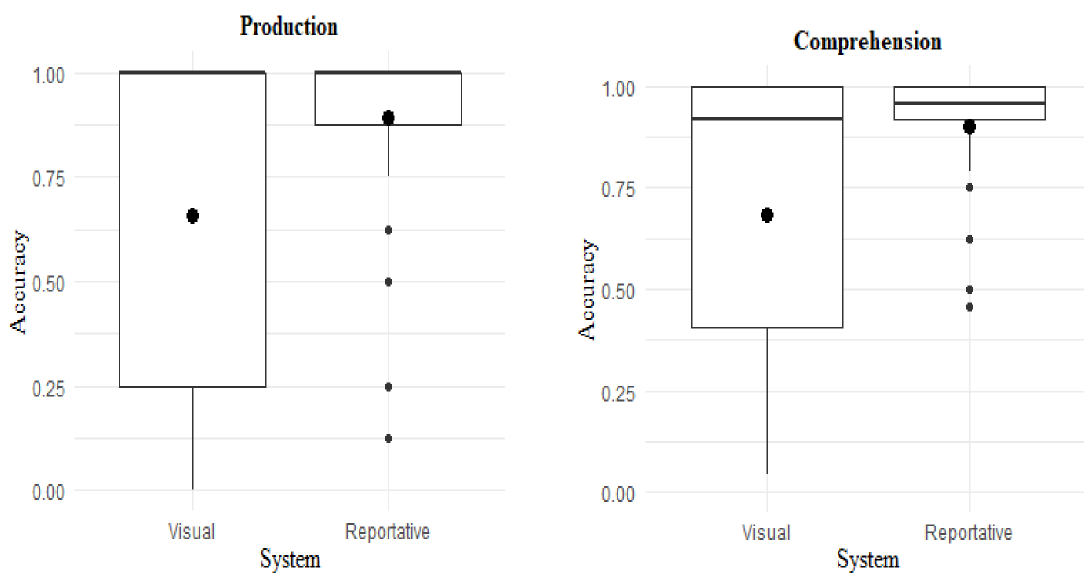

Figure 10. Accuracy score distribution, median (horizontal bar) and mean (dot) in Experiment 6.

Table 9. Parameter estimates for Accuracy in Experiment 6. Significance levels: *p<.05,** p < .01, *** p < .001

Effects	Estimate	SE	z
<u>Production Task</u>			
Intercept	6.63	3.57	1.85
System (Reportative vs. Visual)	-2.50	2.70	-0.92
<u>Comprehension Task</u>			
Intercept	2.91	0.40	7.15 ***
System (Reportative vs. Visual)	-2.07	0.73	-2.79 **

The debriefing data provide additional insight into how the frame was interpreted. Of the 27 participants exposed to the Visual System, 12 correctly associated the presence of the frame with the main character's visual access to what happened: 9 of those participants associated the frame with the Speaker watching the event and 3 specifically mentioned that the character "wasn't blinded by the other girl" or "her eyes are not covered". The remaining 15 participants gave erroneous answers, either associating the frame with the opposite (Reportative) access (n = 7) or giving irrelevant responses (n = 8). Turning to the 32 participants that were exposed to the Reportative System, 26 gave correct responses, mentioning that the speaker had her eyes covered (n=10) or the speaker had her eyes covered and the third character told her what happened (n=16). The remaining participants (n=6) did not provide any evidential conjectures.

Lastly, we compared the results obtained from Experiments 4 and 6 to see whether the learnability patterns differed depending on whether participants tracked linguistic or non-linguistic ways of marking information sources. We ran a mixed-effects logistic regression model for each task (Production and Comprehension) using System (Reportative, Visual) and Experiment (1 vs. 3) as our fixed predictors with Accuracy as our dependent variable. Additionally, our models included random intercepts for Items and Participants. Overall results and parameter estimates for each task can be seen in Table 10.

For Production, both System (Reportative, Visual) (χ^2 = 16.70, *p*< .0001) and Experiment (Language, Symbol) (χ^2 = 5.06, *p*= .02) significantly improved model fit but the interaction between System and Experiment did not (χ^2 = 0.40, *p*= .52). These results suggest that participants performed better when exposed to the Reportative System across both experiments (M_{Rep} = 0.85, M_{Vis} = 0.63), and when the target information source was marked with a non-linguistic Frame compared to an Artificial Language marker (M_{Lang} = 0.70, M_{Symbol} = 0.77). For Comprehension, only the System factor improved the model fit (χ^2 = 22.87, *p*< .0001); there was no significant effect for the factor Experiment (χ^2 = 1.05, *p*= .30), nor for the interaction of System and Experiment (χ^2 = 1.25, *p*= .20).

Table 10. Parameter estimates for Accuracy in the Production and Comprehension tasks of Experiments 4 and 6. Significance levels: ** p < .01, *** p < .001

Effects	Estimate	SE	z	
<u>Production Task</u>				
Intercept	2.14	0.29	7.30	***
System (Reportative vs. Visual)	-2.08	0.50	-4.11	***
Experiment (1 vs. 3)	1.14	0.50	2.82	*
System* Experiment	-0.78	0.97	-0.80	
<u>Comprehension Task</u>				
Intercept	2.78	0.33	8.43	***
System (Reportative vs. Visual)	-2.86	0.59	-4.83	***
Experiment (1 vs. 3)	0.49	0.57	0.86	
System* Experiment	1.45	1.16	1.25	

3.3.3. Discussion

Experiment 3 revealed a learnability pattern strongly resembling that of Experiment 4: participants were more accurate in learning to associate a frame with a character's Reportative access compared to Visual access when they had to judge the appropriateness of the frame that was given for a character's Reportative or Visual access to an event (Comprehension task). When participants needed to specify themselves which events should be marked by the frame (Production task), we got a numerical asymmetry in the same direction that was not, however, statistically significant. Importantly, as in Experiment 4, despite its ubiquity and naturalness as a conceptual category, visual experience was not preferentially marked as an information source even when marked through visual (and not verbal/morphological) means. This general pattern of results is consistent with a pragmatic account of the learnability of evidential distinctions.

How should we explain the precise shape of the current Production results? Notice that, for the Production task, participants had to indicate the presence or absence of a target category by actively choosing whether a frame should be inserted or omitted, thereby becoming active communicators themselves. This step involved adopting a different perspective on the event compared to the character in the videos who gained access to the event (and later talked about it). It is possible that the perspective switch was responsible for the noisier Production data in Experiment 6 and the fact that performance on this task was lower compared to its linguistic version in Experiment 4 (where only the speaker's perspective had to be tracked and adopted by the participant).

In support of this line of reasoning, the Comprehension task that involved no such perspective shift yielded less noisy data (and performance was similar to Experiment 4).

3.4. General Discussion

3.4.1. Pragmatics and the learnability of evidential systems

Within cognitive science it has often been assumed that highly frequent cross-linguistic semantic distinctions should be easier to learn than less frequent ones because they correspond to more natural concepts (e.g., TPH; Gentner & Bowerman, 2009). However, both the scope and potency of this generalization have remained unclear, and empirical work on the learnability of semantics has, in general, been limited (Xu et al., 2013; Carstensen et al., 2015; Maldonado & Culbertson, 2020; Saratsli et al., 2020). Here we explored an additional, or alternative, hypothesis, according to which learnability asymmetries (and the associated typological prevalence patterns) in some cases have pragmatic, as opposed to conceptual, origins.

We focused on evidentiality, since prior work comparing the learnability of multiple novel evidential systems has shown that systems encoding visual access to information – the least frequently marked information source cross-linguistically – are also the hardest to acquire by adult learners (Saratsli et al., 2020). Even though this finding confirms the hypothesis that cross-linguistic prevalence and learnability of semantic distinctions are linked, it presents a puzzle for the dominant view that attributes both prevalence and learnability to the conceptual naturalness of the underlying distinctions (e.g., Gentner & Bowerman, 2009) since visual perception is a concept available early in development and lexically expressed across different languages (Viberg, 1984; Pratt & Bryant, 1990; Ozturk & Papafragou, 2016; San Roque et al., 2018).

Across three experiments, we tested an alternative hypothesis on which the learning bias against visual evidentials is driven by pragmatic factors: learning to mark an indirect, potentially unreliable information source might be easier than learning to mark a source directly associated with perceptual experience and a lower likelihood to be false. In Experiment 1, we compared the ease of acquisition of two miniature evidential systems marking a single information source (Visual or Reportative) by adult native speakers of English whose language lacks evidential morphology. This very simple design removed some of the interpretive possibilities in prior, more complex Artificial Language Learning paradigms used to study the learnability of evidential systems (Saratsli et al., 2020). The pragmatic hypothesis predicted that, even within this very simple design, the Reportative should be the most learnable of the two systems. This prediction was borne out in our data. In Experiment 2, we replicated this paradigm but informed participants that the novel markers they were exposed to had evidential meanings. The pragmatic hypothesis predicted that, even when participants' conjectures were limited to the conceptual domain of evidentiality, pragmatic factors should still offer an advantage to the Reportative compared to the Visual system; again, this prediction was confirmed, with results largely replicating Experiment 1. In Experiment 3, we used the same paradigm except that the participants' task was to identify the kind of event that was marked by a red frame (the presence of the frame tracked either visual or reportative evidence within an event). The preference for the Reportative system persisted (albeit partially), thereby confirming the expectation that even if a speaker's information sources are selectively encoded in non-linguistic ways, reportative sources are still more marked.

Together our data offer evidence for the conclusion that highly frequent semantic distinctions are more learnable than less frequent ones (see Gentner & Bowerman, 2009; Saratsli et al., 2020; for distinctions in syntax, phonology and morphology, see Christiansen, 2000; Newport & Aslin, 2004; Hudson Kam & Newport, 2005, 2009; Seidl & Buckley, 2005; Wilson, 2006; Thompson & Newport, 2007; Wonnacott et al., 2008; Finley & Badecker, 2008; Tily et al., 2011; Merkx et al., 2011; Culbertson, 2012; Culbertson & Smolensky, 2012; Culbertson et al., 2012; Fedzechkina, et al., 2012; Tabullo et al., 2012). Most importantly for present purposes, our data support the assumption that both learnability biases in the semantic domain and the typological frequency of the corresponding semantic distinctions may in some cases have pragmatic – as opposed to conceptual - roots.

Specifically, these results suggest that non-perceptual sources in our data are selectively marked in systems that have a single evidential morpheme because they are informative, i.e., they represent a departure from the primacy of perception as an information source (Grice, 1989; Barnard, Rosen & Matthews, 2017). First-hand information that derives from visual experience is reliable enough to make marking this type of information source of no pragmatic (i.e., communicative) value and hence, more difficult to learn. This is because visual perception as an information source is considered to be in direct correspondence with reality and to have a privileged status as a way of accessing information (Dancy, 1985; Papafragou et al., 2007; Matsui & Fitneva, 2009; Aikhenvald, 2018; Wiemer, 2018). In contrast, indirect information sources, with reported information being the most indirect source category, are less reliable, as they diverge from one's perceptual experience and call for the employment of epistemic vigilance (Sperber et al., 2010). As a result, marking those sources can signify doubt and lack of direct evidence for an event, a pragmatically noteworthy property in communication (Speas, 2018; Wiemer, 2018).

Our data have implications for the way children learn evidential systems. Prior studies indicate that the acquisition of evidentiality is challenging cross-linguistically (Aksu-Koç, 1988; Lee & Law, 2000; Aksu-Koç & Alici, 2000; Matsui, Miura & McCagg, 2006; Matsui, Yamamoto, & McCagg, 2006; Papafragou et al., 2007; Aksu-Koç, Ögel-Balaban & Alp, 2009; de Villiers, Garfield, Gernet-Girard, Roeper, & Speas, 2009; Winans, Hyams, Rett, & Kalin, 2014; Ozturk & Papafragou, 2016; Uzundag, Taşçı, Küntay, & Aksu-Koç, 2018; Ünal & Papafragou, 2018; Fitneva, 2008, 2018). These studies have often hypothesized that part of the difficulty in acquiring evidential distinctions lies in the subtle and abstract conceptual presuppositions of evidential meanings. The present data suggest that, independently of conceptual issues, the acquisition of a single evidential poses pragmatic issues for the learner. Those issues need to be addressed directly in future developmental work. An interesting question in this context is whether children treat visual access asymmetrically compared to other information sources. We know that children realize that, other things being equal, visual perception is a highly reliable information source (Pillow, 1989; Pratt & Bryant, 1990; Robinson, Champion, & Mitchell, 1999; Clément, Koenig, & Harris, 2004; Robinson, Haigh, & Nurmsoo, 2008; Ozturk & Papafragou, 2016); moreover, young children prefer interacting with and learning from reliable rather than unreliable speakers (Sabbagh & Baldwin, 2001; Koenig, & Harris, 2005; Jaswal, & Neely, 2006; Mascaro & Sperber, 2009; Koenig, 2012; Jaswal, 2010; Fusaro, Corriveau, & Harris, 2011; Harris, 2012;). Furthermore, we know that adults at least are sensitive to the epistemic meaning of information source marking (e.g., they recognize that indirect sources of information convey uncertainty or reduced reliability; Arslan, 2020; Degen et al., 2019). It remains to be seen how these considerations shape hypotheses about the meaning of unknown evidentials as learners acquire evidential systems in their native language.

3.4.2. Pragmatic vs. conceptual naturalness and the learnability of semantics

Taken most broadly, our data show that certain semantic distinctions are easier to learn (and more prevalent across different languages) not because of the naturalness of the expressed meanings but because of their communicative contribution. On this picture, pragmatic pressures can be viewed as one of the many factors that learners of any age use to map otherwise available concepts onto linguistic forms in their input (Gleitman, 1990). Because pragmatic pressures such as the need to be informative are presumably universal (Grice, 1975; Sperber & Wilson, 1986), they can capture robust and cross-linguistically stable features of both the shape and acquisition of semantic distinctions. This perspective comports with evidence showing that, early on in life, humans selectively comment on whatever they find unexpected and hence informative (Greenfield & Smith, 1976; Greenfield, 1979), and even young children calculate informativeness as they learn and use language (Bannard et al., 2017; Grigoroglou & Papafragou, 2019a, b; Kampa & Papafragou, 2020; Katsos & Bishop, 2011; Frank & Goodman, 2014).

The present perspective generalizes to several other domains beyond evidentiality. Recall that, within the spatial domain. certain spatial expressions are acquired earlier than others cross-linguistically. For instance, spatial terms roughly equivalent to *in*, *on, under* seem to be acquired earlier than *back* and *front* (acquired in that order); other terms appear even later (Johnston & Slobin, 1979). As alluded to in the introduction, these robust learnability differences have traditionally been attributed to conceptual asymmetries (e.g., the basicness of simple topological concepts encoded by *in, on, under* compared to the complexity of projective concepts such as *front/back* that require the calculation of axial information (ibid.). Within the projective class, further conceptual asymmetries have been invoked to explain why *back* appears early in child speech and is used more frequently than *front* (Johnston & Slobin,1979). In contrast to this simple conceptual hypothesis, a wealth of experimental evidence points towards a more complex picture of the acquisition of spatial semantic terms that is sensitive to a multiplicity of options for encoding semantic relations (e.g., Landau et al., 2010). Importantly for present purposes, the strive for informativeness often seems to be at the heart of such cross-linguistic patterns. For instance, even adults prefer to use *back* more than *front* cross-linguistically, and the reasons seem to be related to the informational gains from using *X is to the back of Y* to locate what are typically occluded objects (Grigoroglou, Johanson & Papafragou, 2019). Many other asymmetries in the use and acquisition of spatial prepositions can also be attributed, at least in part, to pragmatic (informativeness) effects (see, e.g., Do, Papafragou & Trueswell, 2020 on the source/goal asymmetry in motion events).

A second example comes from mental verbs such as *think*, *believe* and *know*. Such verbs have traditionally been considered hard to acquire because the underlying abstract concepts are hard to grasp (Huttenlocher, Smiley, & Charney, 1983). Yet, in a set of findings highly reminiscent of the present work, Papafragou, Cassidy and Gleitman (2007) provided evidence that both young children and adults interpret novel verbs as mental verbs when they were presented with a story in which the verbs used described *false* beliefs. For instance, if a story involved deception similarly to the classic Red Riding Hood tale, the novel verb in a sentence such as "Matt gorps that his grandmother is under the covers", was more likely to be interpreted as a mental verb (e.g., *think* or *know*) by both young and more experienced (adult) learners compared to a story where Matt's belief was true. Thus, talking about belief's was expected when these beliefs did not correspond to reality and hence presented a pragmatically notable property (cf. also Hacquard & Lidz, 2019). Similarly, in our current findings, perceptually supported (and presumably true) beliefs were less likely to be highlighted compared to indirectly acquired (and potentially false) beliefs.

Most broadly, our proposal is compatible with evidence pointing to the centrality of pragmatics in language acquisition (for reviews, see Grigoroglou & Papafragou, 2017, in press; Clark & Amaral, 2010; Matthews, 2014). We know that children's acquisition and use of words in several semantic domains including modals (Ozturk & Papafragou, 2015), aspectuals (Papafragou, 2006; Wagner, 2009), quantifiers (Noveck, 2001; Papafragou & Musolino, 2003; Lidz & Musolino, 2006), gradable adjectives (Syrett, Kennedy, & Lidz, 2009), numbers (Musolino, 2004; Barner & Bachrach, 2010), negation (Nordmeyer & Frank, 2014; Reuter, Feiman & Snedeker, 2018), mass/count terms (Srinivasan & Barner, 2016), and color vocabulary (Wagner, Tillman & Barner, 2016), is a function not just of the semantic content of these words (and hence the underlying concepts) but also of their pragmatics. In many of these areas, children seem to acquire new members of the lexicon by making conjectures about new meanings on the basis of contrasts with existing words in the language (see E. Clark, 2010, for a classic formulation of this idea). As in our own experiments, many hypotheses about what a linguistic expression encodes rest on comparisons to other lexical alternatives (or to the absence thereof).

Beyond the developmental field, the present proposal has affinities with research exploring the role of pragmatics in shaping the adult lexicon. For instance, it has been suggested that languages lack words for meanings that are canonically conveyed by stable and ubiquitous pragmatic implicatures (e.g., cross-linguistically, there is no quantifier meaning ‘not all’, presumably because the same meaning is a strong pragmatic inference from the use of *some*; Horn, 1989). This direction has recently led to computational approaches that have linked the organization of the lexicon cross-linguistically to communicative pressures related to efficiency (Gibson et al., 2017; Kemp, Xu & Regier, 2018). This idea of efficiency highlights the fact that language, as a vessel for communication, needs to be both maximally informative but also simple: the intended message should contain all the essential information needed to fully understand what is being communicated but this needs to happen using minimal cognitive resources (Regier, Kemp, & Kay, 2015; Regier, Carstensen, & Kemp, 2016; Gibson et al., 2019; Steinert-Threskeld & Szymanik, 2020).

Chapter 4

ACQUIRING EVIDENTIAL MEANINGS: MAPPING ACROSS DIFFERENT FORMS

So far, our experimental studies have provided evidence that point to that fact that people preferentially learn to mark indirect information sources – mostly third-person reports- through novel morphemes and non-linguistic symbols, guided by the pragmatic pressures that potentially unreliable information sources can carry. However, evidential meanings can be potentially mapped onto different language forms, beyond morphological elements and the question that arises is what factors are at play when learners are in the process of making these form-meaning associations.

Aside from grammaticalized evidentiality systems, there is a variety of verbs and adverbs that can be used as an evidential strategy, expressing speaker's source of information and sometimes also degree of certainty (Aikhenvald, 2007; Aikhenvald & Storch, 2013). Such lexical expressions of information source can be found cross-linguistically, in a wide range of different language families (e.g. European languages: Cornillie, 2007; Diewald & Smirnova, 2010; Aikhenvald, 2014; Boye, 2018; Uralic languages: Skribnik & Kehayon, 2018; Formosan languages: Pan, 2018). Turning specific to English, perception verbs (e.g. see, hear) are prevalent, with verbs encoding visual perception communicating not only a direct sensory experience but also speaker's "internal knowledge" (Squartini, 2018). These meanings and their association with source of knowledge are available early on in development (Pillow, 1989; Pratt & Bryant, 1990; Clément, Koenig, & Harris, 2004; Ozturk & Papafragou, 2016) and are even accessible by blind individuals (Landau & Gleitman, 1985; Bedny et al. 2019). Additionally, there are certain adverbial expressions that indicate speaker's information source, such as *clearly* or *obviously* (e.g. *Obviously*, she is lying) with specific adverbial cases used to explicitly state that the proposition of the speaker's utterance is hearsay (e.g. *allegedly, reportedly*; Urmson, 1963; Palmer, 1986; Ifantidou-Trouki, 1993; Rocci, 2017). Consequently, the question that arises is what learnability patterns could arise if learners were asked to learn novel lexical forms that are meant to express evidential meanings and what type of factors, if any, would influence the form-to-meaning association and the overall learnability of different forms.

When it comes to the acquisition of the grammaticalized evidentiality, crosslinguistic developmental evidence point to the fact that evidential markers are hard to learn (Aksu-Koc & Alici, 2000; Unal & Papafragou, 2016; Fitneva, 2018). Children, even up to the age of 5, face difficulties in differentiating the different types of evidential markers present in a language and then in turn, associating these markers to the appropriate sources (Tibetan: de Villiers, Garfield, Speas, & Roeper, 2007; de Villiers et al., 2009; Japanese: Matsui, Yamamoto, & McCagg, 2006; Turkish: Ozturk &

Papafragou, 2015). Research evidence from Turkish suggests that children's production of evidential morphemes is at chance around the age of 4 and specifically, the use of the indirect markers seems to be inconsistent even at the age of 6 (Aksu-Koç, 1988). Additionally, Korean's children's comprehension of evidential morphemes does not follow a stable developmental trajectory from three to six years of age (Papafragou et al., 2007; Choi, 1995).

According to one proposal, such difficulties derive from the fact that evidential expressions can carry such "epistemic overtones" pointing to a speaker's cognitive state (Wiemer, 2018) and hence, they denote a conceptually complex notion of access to information that is abstract and not directly observable (Aksu & Slobin, 1986; Gopnik and Graf, 1988; Ozturk & Papafragou, 2015). This proposal echoes a similar hypothesis that has been proposed relevant to the acquisition of mental verbs (such as *think*, *believe*). These verbs also express abstract and complex mental states and indirect modes of knowledge that cannot be easily extracted from observable events and constitute a late developmental achievement, presumably deriving from children's conceptual difficulties (Bartsch & Wellman, 1995; Gopnik & Meltzoff, 1997; cf. Gillette et al., 1999; Papafragou, et al., 2007).

Nonetheless, this proposal cannot account for the fact that information source concepts and the ability to monitor those sources seem to be available quite early, prior to the acquisition of evidential language (Matsui, Miura, & McCagg, 2006; Papafragou, Li, et al., 2007; Ozturk & Papafragou, 2015; Unal & Papafragou, 2020). Evidence from source monitor tasks show that three year old children associate visual perception with knowledge source and preferably prefer to believe an adult informant that has visual access to an event compared to one that has not (Pillow, 1989; Pratt & Bryant, 1990; Robinson, Champion, & Mitchell, 1998). Children also seem to be able to distinguish different sources of information: they can explain the different ways they get to know about something (Gopnik & Graf, 1988;O'Neill & Astington, 1990; O'Neill & Gopnik, 1991; Woolley & Bruell, 1996; O'Neill & Chong, 2001) and they also seem to disprefer adults' verbal reports over seeing something for themselves (Clément, Koenig, & Harris, 2004; Robinson, Haigh, & Nurmsoo, 2008). This ability to monitor different sources of information certainly continues to develop throughout children's early years but delays in the acquisition of linguistic evidential expressions do not seem to be deriving from conceptual difficulties related to information source concepts.

On a second proposal, the delayed developmental trajectory is derived from difficulties in extracting the correspondence between information source concepts and the target linguistic forms (Matsui et. al., 2006; Ozturk & Papafragou, 2015). Linguistic, structural patterns can act as cues to learners for helping them deduce the association between the target concepts and forms (Gleitman, 1990; Gleitman, et. al., 2005). Consequently, difficulties in using such linguistic cues to parse what is being referred to, can lead to delays in the acquisition of the target form-to-meaning mappings, even if they have already developed the underlying information source concepts. As shown by the results of the human simulation paradigm in which participants need to guess the meaning of different words, verbs that expressed a more mentalistic and abstract concept (e.g. *think*, *want*) were harder to be gleaned out merely from a situational context with no other linguistic constraint compared to verbs with more concrete meanings (e.g. *push*) or nouns (Gillette, Gleitman, Gleitman & Lederer, 1999; Snedeker & Gleitman, 2004; Gleitman, Cassidy, Nappa, Papafragou & Trueswell, 2005). Additionally, research evidence points to the fact that syntactic constraints (often in combination with pragmatic cues relevant to speaker's intention) assist in figuring out the semantics of such complex, mentalistic concepts (Gleitman, 1990, Gleitman, et. al., 2005; Papafragou, Cassidy and Gleitman, 2006; Hacquard and Lidz, 2019).

Consequently, our next goal was to assess the *unique* effect that mapping factors may have on the learnability of evidential meanings when these are expressed through different linguistic forms. To ensure that any differences that arise are solely derived from the effects of the mapping process, the linguistic forms that learners would need to learn should express the same information source concepts. Using a similar Artificial Language Learning paradigm as before, adult learners were exposed to events in which the speaker gets access to information through visual perception (*Visual access*) or through another person's reports (*Reportative* access) and only one access type was marked each time through different, novel linguistic forms: a verb, an adverb or a morpheme. Therefore, each linguistic form is used to express the same underlying concept (Visual or Reportative access to an event) and any learnability differences can be tied with the linguistic constraints that unique for each.

In more detail, each participant would be exposed to only one of the linguistic forms: participants exposed to a novel verb as the marking tool would see the speaker describing the sentence a sentence construction similar to this: *"I gorp* she copied the drawing". Participants exposed to the adverb as the evidential marker would see a sentence that would start with a novel adverbial such as: "Gorpingly, she copied the drawing" while those exposed to the morpheme as the evidential marker would see that marker appear as a verb final suffix: "She copied*gorp* the drawing".

The question that arises therefore is whether one of these forms invites more direct form-meaning associations, leading in turn, to higher learning accuracy. The prediction set forward here is that the different syntactic properties that these linguistic forms bear will affect participants' learnability. For the case of the novel verb, the presence of the sentential complement (*"I gorp* [she copied the drawing]") can guide learners to a potential meaning interpretation such as *I see* or *I hear* depending on the access type that is getting marked. However, the syntactic properties for the novel adverb and the novel morpheme do not offer any stable syntactic anchors. The adverbial syntactic position in the beginning of the sentence does not necessarily limit the syntactic scope of the expression as the adverbial expression could modify either the proposition as a whole (pointing to an evidential meaning, e.g. *allegedly*) or the verb phrase of the proposition (pointing to a non-evidential meaning, e.g. *carefully*). For the morpheme, the potential meaning-form association is even more opaque as there is no readily available syntactic cue that can be extracted and utilized for meaning disambiguation. There was an additional condition in which participants were exposed to a red frame being used as a means of marking the target access type. This condition served as a learning baseline, with the intention to compare in what way mapping these meanings on linguistic forms instead, influences participants' learning accuracy.

Additionally, there is an additional prediction that is related to the pragmatic nature of the two information sources that are being encoded through these forms. As we have seen so far, a pragmatic bias has led participants to prioritizing marking reportative over visual meanings due to the fact that visual experience is more ubiquitous and reliable, and hence less informative and pragmatically noteworthy, making it less likely to be marked (Papafragou et al., 2007; Matsui & Fitneva, 2009; Aikhenvald, 2018; Wiemer, 2018). Consequently, the prediction is that this pragmatic bias to mark reportative information sources should persist in this set of studies as well, both for linguistic and non-linguistic forms.

4.1. Experiment 7

4.1.1. Methods

4.1.1.1. Participants

We recruited 280 participants between the ages of 18 and 70 through Amazon Mechanical Turk (https://www.mturk.com/). There were 78 participants were exposed to a novel verb, 72 participants to the novel adverb, 68 participants to the novel morpheme and 80 participants to the non-linguistic symbol. For each form, participants were equally divided and randomly assigned to one of the information access types (Reportative or Visual access). All participants were native speakers of English with no one speaking an additional language at home or a language that included grammatical evidentials.

4.1.1.2. Stimuli

There were 42 videos, presented in two versions, with each version displaying either Visual or Reportative access to information. The content of the videos was identical to those reported in the experimental procedures in the preceding chapter: each video depicted one of the characters performing an action ("the Agent") while a second character either fully witnessed this action (Visual Access version) or was informed about what happened without having any type of visual experience of the event (Reportative Access version) and described what happened (hence "the Speaker"). A third character, sitting behind the two other characters, either allowed the Speaker to watch what happened or blocked her visual access throughout the event.

At the end of each video, a speech bubble appeared with the character's description of the event: the speaker's sentence was in English but depending on the experimental condition that participants were exposed to, there was either a novel linguistic form or a non-linguistic symbol, intended to mark the target information access type. Starting with the linguistic forms, participants would either see a sentence with a novel verb ("*I gorp she lit the lamp*"), a novel adverb ("*Gorpingly, she lit the lamp*") or a novel morpheme ("*She litgorp the lamp*") appearing in the beginning of the sentence. For the case of the non-linguistic symbol, the bubble would contain an English sentence (e.g. *She lit the lamp*) but for the depending on the target access type that was being marked, the videos displaying that access type would be surrounded by a red frame throughout the event. Therefore, in our experimental paradigm participants are either assigned to either a

Visual or a Reportative system depending on which access type is getting marked by linguistic or non-linguistic form.

Within each system, the events corresponding to the other type of access were described using plain (unmarked) sentences. Additionally, each participant would be exposed to one form being used to mark the target access type (a novel verb, a novel adverb, a novel morpheme or a red frame). Therefore, our design has two between-subject factors: the Evidentiality System that participants are randomly assigned to (Visual or Reportative) and the Form that is being used to mark the events with the target access type (verb, adverb, morpheme or frame).

As before, our experimental paradigm consisted of a Training Phase a subsequent Testing Phase, in which participants' ability to use (Production Task) and understand when to use the target form (Comprehension Task) was assessed. Each event was filmed in two versions depending on the Speaker's Access to the event (Visual or Reportative), which resulted in having two counterbalancing lists by rotating the Speaker's access to the event (i.e., if for an event in list 1 showed Speaker had Visual Access, for the **same** event in list 2 the Speaker would have Reportative access). For each list, we created two different presentation orders for each evidential system, resulting in four different randomized presentation lists per evidential system. Participants are randomly assigned to one of the four corresponding presentation lists of that evidential system. The same counterbalancing and presentation lists were used for each form type (verb, adverb, morpheme, frame).

4.1.1.3. Procedure

Our experimental sessions were administered through Ibex PennController (Zehr & Schwarz, 2018) and similarly to our previous experiments, participants in the beginning were introduced to the setting of the events and to what the role of each character appearing in the videos would be in an effort to familiarize them with the content of the stimuli before the actual videos were displayed. Subsequently, moving on to the Training Phase of the experiment, participants were provided with information about the marker they would see for some of the events and what they needed to do. More specifically, for the linguistic forms (verb, adverb and morpheme), participants were told that:

> "The characters will be speaking an alien language. This language shares some words with English but one difference is that it includes a special word, "gorp"/ "gorpingly".
> You will have to pay attention to when "gorp"/ "gorpingly" appears in order to try and figure out what it means."

For the frame condition, the instructions slightly changed as follows:

> "Some of the events will have a red frame. You will have to pay attention to when the red frame appears in order to try and figure out what kind of event gets this frame. You will go through several events to figure this out."

Following that, across all forms and evidential systems, participants were informed that there would be a second part in which their understanding of where the marker or symbol appeared based on its meaning would be assessed. The Training Phase included 10 videos in total, 5 videos for each Access type (Visual, Reportative).

After the Training Phase was completed, participants proceeded first with the Production and later with the Comprehension Task. For the Production Task, there were

8 new videos, 4 videos for each type of Access and the structure of these videos was identical to the videos shown in the Training Phase. However, this time, there was no description or red frame appearing. For the linguistic forms (verb, adverb and morpheme), when the speech bubble appeared at the end of the video, participants were presented with the verb and the noun relevant to the event and they were asked “to type in whether the speaker should describe what happened with either using or not using the novel word” (see Figure X) depending on whether they thought the marker was needed for each given trial considering what they figured out from the Training Phase. Participants were given time to thoroughly read and understand the instructions and they were also shown a sentence example that was not part of the stimuli so as to ensure they understood what they needed to do and what their answer should look like.

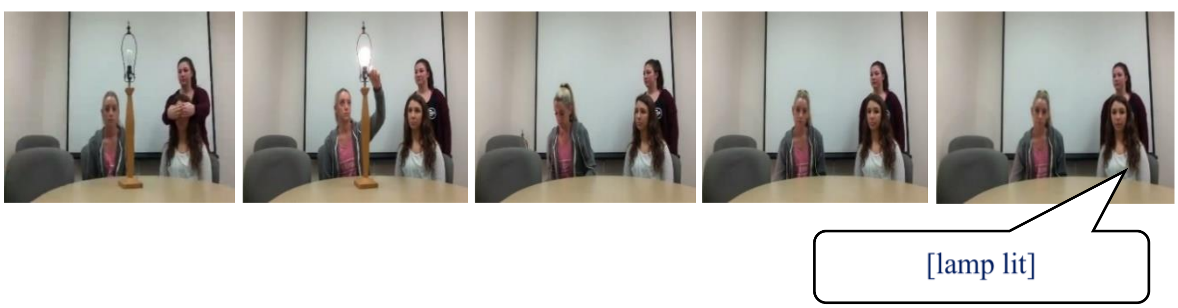

Figure 11. In the Production task, the speech bubble at the end of the video contained the verb and the noun relevant to what they watched.
They had to decide whether they should write a sentence that contained the marker or not depending on the access type

For instance, for the event provided in Figure X, the participant group that was exposed to the verb form as the means of marking the speaker’s access to information, they would have to decide whether they need to type in an unmarked sentence (“She lit the lamp”) or with the marker (“I gorp she lit the lamp”) depending on what they learned in the first part of the experiment (i.e. whether the access type shown in the video should be marked or not).

For the participant group for which the access type was marked through a red frame, the instructions and requested responses were a little different. Participants were told that this time no red frame will appear and they will have to decide whether each video they watched should have gotten a red frame or not. More specifically, participants were told that:

> “Almost everything will be the same as before BUT this time no frame is included. For each event, please type “Frame” if you think that they event should have a red frame or “No frame” if you think that no frame is needed.”

Turning to the Comprehension task, we used 24 videos, 12 per each Access type. For the participant groups exposed to the linguistic markers, for half of the events, participants were told that the Speaker erroneously used the linguistic marker, either by omitting the marker when it should have been used or by using it for the wrong type of access. For the remaining half of the events, the use of the marker was correct. Participants had to type in either ‘yes’ or ‘no’ depending on whether they thought the character was correctly using the marker or not.

For the frame condition, participants were told that some of the events would have a red frame when it is not appropriate, or they would be missing the red frame when it was needed. So, what participants had to do was to write ‘yes’ or ‘no’ depending on whether the frame was correctly missing or included. After completing the Comprehension Task, participants were transferred to a new page where they were asked to provide a brief description of what they thought the marker or the symbol meant and when it was used or appeared respectively.

4.1.2. Results

4.1.2.1. Comparison across linguistic forms

Participants' responses were coded as a binary outcome variable (using 1 for a correct response or 0 for an incorrect response), reflecting their accuracy in each task. Summary data for learning accuracy across the different forms can be seen in Figure 12 while Figure 13 displays learning accuracy across forms and evidential systems. A logistic mixed-effects modeling was used (Baayen, 2008; Baayen, Davidson, & Bates, 2008), implementing the generalized binomial linear mixed effects (glmer) function of the lme4 package since the target data contained categorical variables (Bates, Maechler, Bolker, & Walker, 2015). *R Project for Statistical Computing* was used for the analysis (R Development Core Team, 2018).

For the current analysis, the aim has been to assess any learning differences when participants needed to map evidential meanings across different forms. Consequently, for the purposes of this analysis, participants' accuracy scores for each task (Production, Comprehension) were combined. The mixed-effects model included participants' accuracy as the dependent variable and the type of evidential System (Visual, Reportative) and Linguistic Form (verb, adverb, morpheme) as the fixed predictors along with their interaction. Random intercepts were also included for Participants and Items. The fixed predictor of System had two levels and we assessed its effect with one planned comparison between the Reportative and the Visual systems (contrast coding: -.50,.50). For the Form predictor, there were three levels (verb, adverb and morpheme) and consequently, two planned comparisons were included comparing the morpheme condition against the other two conditions: morpheme against verb (contrast coding:.50, 0, -.50) and morpheme against adverb (contrast coding: 0, .50, -50).

Including both fixed predictors in the model significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood (χ^2 = 73.85, p< .00001) compared to when including their interaction (χ^2 =.45, p= .79). Participants had higher accuracy rates in the verb condition compared to the morpheme condition (M_{Verb}, = 0.93, $M_{Morpheme}$ = 0.61, p< .00001) but there was no difference in the accuracy rates for adverb and morpheme (M_{Adverb}, = 0.81, $M_{Morpheme}$ = 0.61, p= .79). There are also higher accuracy rates for the Reportative compared to the Visual System (M_{Rep}, = 0.84, M_{Vis} = 0.75, p= .0002). The parameter estimates of the model can be found in Table 11.

Table 11. Parameter estimates for Accuracy in Experiment 7. Significance levels: * p < .05, ** p < .01, *** p < .001

Effects	Estimate	SE	z
Intercept	2.42	0.15	15.51 ***
System (Reportative vs. Visual)	-1.00	0.27	-3.67 ***
Verb vs Morpheme	3.03	0.39	-7.62 ***
Adverb vs Morpheme	0.09	0.38	-0.25

4.1.2.2. Comparison across linguistic and non-linguistic forms

Next, the goal was to compare participants' learning accuracy in each of the linguistic forms against their learning accuracy for the non-linguistic symbol (the red frame). The model was similar to the one used in the previous analysis: we included participants' accuracy as the dependent variable and the main effects and interaction of the type of evidential System (Visual, Reportative) and Form (verb, adverb, morpheme, frame) as the fixed predictors along with random intercepts for Participants and Items. System effect was assessed by one planned comparison (contrast coding: -.50,.50). For the Form predictor, three planned comparisons were included: comparing accuracy for verbs against accuracy for frame (contrast coding: .50, 0, 0, -50), accuracy for adverbs against accuracy for frame (contrast coding: 0, .50, 0, -50) and lastly, accuracy for morphemes against accuracy for frame (contrast coding: .0, 0, .50, -50).

Including both fixed predictors in the model significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood ($\chi^2 = 2.28$, $p < .00001$). There was no significant improvement when adding the interaction between the two predictors ($\chi^2 = 2.28$, $p = .51$). Participants had higher accuracy rates in the verb condition compared to the frame condition (M_{Verb}, = 0.93, M_{Frame} = 0.78, , $p < .00001$) with no difference in the accuracy rates between the adverb and frame conditions (M_{Adverb}, = 0.81, M_{Frame}= 0.78, $p = .90$). Additionally, there is a significant difference in the accuracy rates for morpheme and frame conditions: morpheme conditions yields a significantly lower accuracy rate compared to the accuracy rates for the frame condition ($M_{Morpheme}$, = 0.61, M_{Frame}= 0.78, $p < .00001$). Turning to the effect of System, participants performed better in the Reportative compared to the Visual System (M_{Rep}, = 0.85, M_{Vis} = 0.73, $p < .00001$). The parameter estimates of the model can be found in Table 12.

Table 12. Parameter estimates for Accuracy across linguistic and non-linguistic forms. Significance levels: * p < .05, ** p < .01, *** p < .001

Effects	Estimate	SE	z
Intercept	2.53	0.15	16.59 ***
System (Reportative vs. Visual)	-1.29	0.27	-4.72 ***
Verb vs Frame	-3.12	0.48	-6.48 ***
Adverb vs Frame	3.28	0.46	7.07
Morpheme vs Frame	-0.05	0.46	-0.11 ***

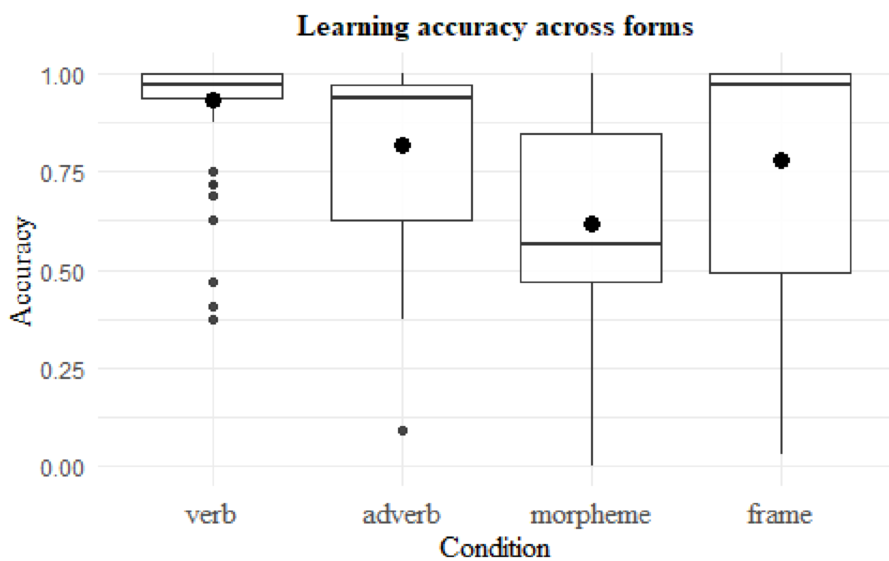

Figure 12. Accuracy score distribution, median (horizontal bar) and mean (dot) in Experiment 7.

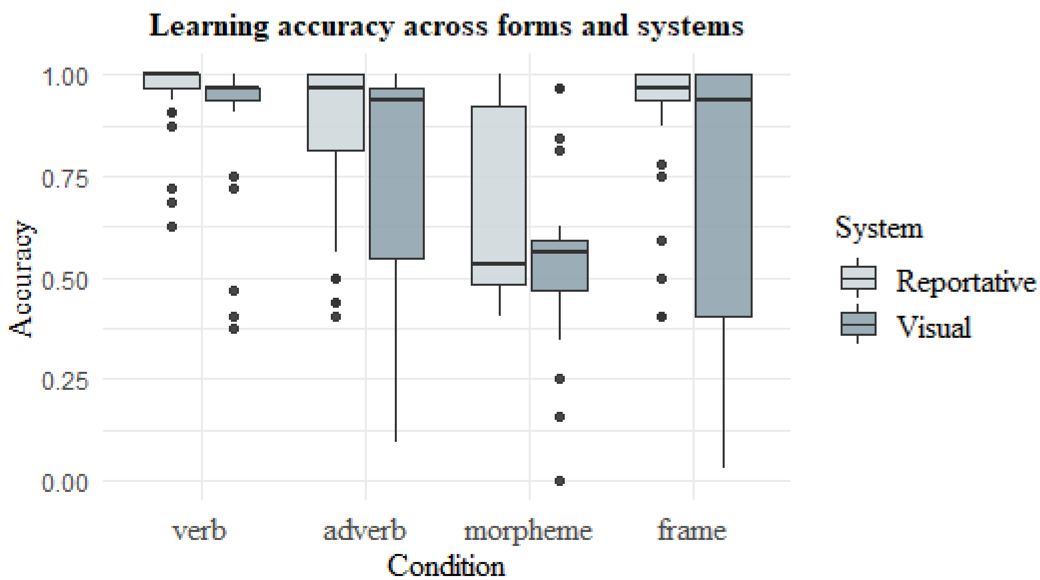

Figure 13. Accuracy score distribution and median (horizontal bar) in Experiment 7.

At the end of the experimental session, participants were asked what they thought the novel form meant and how it was used. Within the verb condition, out of the 40 participants that were exposed to the Reportative System, 15 reported the form meaning that the speaker "*was told*" or that "*she heard about what happened by someone else*". Another 16 participants connected the verb form with a mentalistic meaning (*think*/*believe*). The rest 9 participants provided answers that were irrelevant to evidential or similar meanings. Out of the 40 participants that were exposed to the novel verb form marking Visual Access, 32 reported that the novel verb meant that that the speaker "*saw the action*" herself while 2 participants associated with the verb "*know*".

Out of the 37 participants that were exposed to the adverb form marking Reportative Access, 16 participants mentioned that it means "*not see*", 9 of them that it means that the speaker "*was told*" about the event while 2 participants connected the adverb form with mentalistic verbs "*think*" and "*guess*". There was also an additional participant that connected the novel adverb with the action of whispering. The rest 9 participants did not successfully associate the adverb form with relevant meanings. For the 35 participants that were exposed to the adverb form marking Visual Access, 20 participants connected with the adverb with the meaning of "*visibly*" or "*saw*" while one participant specifically connected the adverb form with "*certainly*". There rest 14 participants provided irrelevant meaning associations.

For the morpheme condition, out of the 31 participants exposed to the Reportative Access, 6 participants associated the novel morpheme with the speaking being told about the event while another 5 participants connected the morpheme with the inability to see ("*not see*"). the rest 20 participants provided irrelevant meanings. For the Visual Access system, out of the 37 participants, only 4 associated the morpheme with the speaker being able to see what is happening while the rest 33 provided irrelevant meanings.

Lastly, turning to the non-linguistic condition where the red frame is used as a marker, out of the 40 participants exposed to the Reportative Access, 15 participants associated the red frame with the speaker being told about the event while another 9 participants associated it with the speaker not being able to see throughout the event. The rest 16 participants made irrelevant meanings associations. For the 40 participants that were exposed to the red frame as a marker for Visual Access, 12 participants connected the frame with meanings such as "*see*", "*witness*" and "*watch*". 7 participants mistakenly associated the red frame with meanings that were pointing to the opposite type of Access (i.e. Reportative Access). The rest 21 participants provided irrelevant meaning explanations.

4.1.3. Discussion

In this experiment, we wanted to assess the unique effect of mapping factors on the learnability of evidential meanings when these are mapped on different linguistic and non-linguistic forms. The first prediction was that the learnability of the different linguistic forms (mapping evidential meanings on novel verbs, adverbs and morphemes) will be affected by the syntactic properties of each form, with the sentential complementation structure of the verb form providing the strongest syntactic clue for the form-meanings association. This prediction is actually borne out: participants learn more easily to map evidential meanings onto novel verbs compared to when they learn to map these meanings onto morphemes. This indicates that the syntactic anchor of sentential complementation that is present for the novel verb is guiding learners to the target meaning associations. On the other hand, for the case of the verb-final morpheme, the opaqueness of the correspondence between the form and the evidential meaning does not allow participants to successfully encode and learn the target form-meaning association.

Additionally, there is no difference in the learnability patterns for the morpheme and the adverb forms: for both, participants seem to face difficulties in deciphering the target meaning which in turn leads to low learning accuracy rates. Similarly to morphemes, the adverbial expression's position in the beginning of the sentence does not offer a syntactic anchor that can confidently lead learners to evidential meanings. This syntactic position could allow the adverbial expression to modify either the proposition which in this case it would be guiding learners to an evidential meaning, (e.g. *allegedly*) but it could be also modifying the verb phrase of the proposition which would in turn point learners to a non-evidential meaning (e.g. *carefully*).

Furthermore, comparing participants' learnability rates for each linguistic form against the learnability rates observed for when information access was marked with a non-linguistic symbol, revealed an interesting pattern. Participants seem to acquire more easily a novel verb form marking evidential meanings than when these are marked through a non-linguistic symbol. This learnability difference can be attributed to the syntactic properties that the verb form creates that enable the meaning-form association while a non-linguistic symbol obscures such an association. However, despite the obscurity that this symbol creates, learners seem to have an even lower learnability rate when these evidential meanings are marked through a verb-final morpheme. This confirms the fact that the verb-final morpheme creates an opaque correspondence between this linguistic marker and the evidential meanings, such that it has a detrimental effect on their learning. On the other hand, there seems to be no learning difference for when the target meanings are expressed through an adverbial expression or a non-linguistic symbol. This potentially derives from the fact that, in a similar way with the red frame, the adverbial expression does invite some type of form-meaning mapping but this mapping is not definitively clear and therefore participants do not create a strong association with evidential meanings.

The second prediction set forward in this experiment was the fact that this pragmatic bias to mark indirect, potentially unreliable information sources would lead participants to learn to mark a reportative information source easier than marking visual experience of an event. This is precisely the pattern observed in the current learnability data: participants learned to mark reportative information access meanings more easily across all different forms (both linguistic and non-linguistic), confirming learners' tendency to drive to mark what is unexpected or pragmatically notable.

Overall, the experimental evidence points to the fact that both the the syntactic structure and the pragmatic properties can facilitate the extraction of commonalities across evidential meanings and benefit (or even negatively affect) the learning of form-to-meaning mappings.

4.2. Experiment 8

The next goal was to assess participants' ability to map evidential meanings across different forms and how their learning accuracy would differ if they were provided with a cue pointing towards the intended form-meaning association. The presence of such cue would eliminate any confounding factor relevant to confusion or misinterpretation on the participants' side.

4.2.1. Methods

4.2.1.1. Participants

For this experiment, there were 293 participants recruited in total, between the ages of 18 and 70 through Amazon Mechanical Turk (https://www.mturk.com/). 65 of these participants were exposed to a novel verb, 78 participants to the novel adverb and 70 participants to the novel morpheme. The remaining 80 participants were exposed to the non-linguistic symbol. For each form, participants were equally and randomly assigned to one of the information access types (Reportative or Visual access). They were native speakers of English and no participant reported speaking an additional language at home or a language that included grammatical evidentials.

4.2.1.2. Stimuli and Procedure

All stimuli (including videos, presentation and counterbalancing lists) remained identical to the previous experiment.

For the linguistic forms, the instructions participants received prior to the Training Phase were modified in the following way (marked in bold font):

> "The characters will be speaking an alien language. This language shares some words with English but one difference is that it includes a special word, "gorp"/ "gorpingly".
> **The speaker uses this word to communicate how she got to know what happened**"

For the case of the non-linguistic marker (the red frame), the modified instructions were as follows(once again, changes indicated in bold font):

> "Some of the events will have a red frame. **This frame is connected with how the speaker got to know what happened.** You will have to pay attention to when the red frame appears in order to try and figure out what kind of event gets this frame."

The Training Phase and the Testing Phase (both the Production and the Comprehension task) remained identical to those of Experiment 7.

4.2.2. Results

4.2.2.1. Comparison across linguistic forms

Similarly to the previous analysis, the first aim was to compare the learning accuracy across the different linguistic forms (verb, adverb, morpheme). Participants' responses were once again coded as a binary outcome variable (using 1 for a correct response or 0 for an incorrect response. Consequently, a logistic mixed-effects modeling was used (Baayen, 2008; Baayen, Davidson, & Bates, 2008), implementing the generalized binomial linear mixed effects (glmer) function of the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). The analysis was performed using *R Project for Statistical Computing* (R Development Core Team, 2018). Summary data for learning accuracy across forms and across forms and evidential systems can be found in Figures X and Y respectively.

Participants' accuracy was included as the dependent variable and the type of evidential System (Visual, Reportative) and Linguistic Form (verb, adverb, morpheme) as the fixed predictors along with their interaction. We also included random intercepts for Participants and Items. The fixed predictor of System had two levels and we assessed its effect with one planned comparison between the Reportative and the Visual systems (contrast coding: -.50,.50). For Linguistic Form, there were two planned comparisons, each comparing the morpheme condition against the other two conditions: morpheme against verb (contrast coding: .50, 0, -.50) and morpheme against adverb (contrast coding: 0, .50, -50).

Including both fixed predictors in the model significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood ($\chi^2 = 25.17$, $p < .00001$). Including the interaction of the two predictors did not significantly improve model fit ($\chi^2 = 2.38$, $p = .30$). Participants had higher accuracy rates in the verb condition compared to the morpheme condition (M_{Verb}, = 0.90, $M_{Morpheme}$ = 0.69, $p < .00001$). However, no significant difference was observed in the accuracy rates for adverb and morpheme (M_{Adverb}, = 0.78, $M_{Morpheme}$ = 0.69, $p = .79$). There are also higher accuracy rates for the Reportative compared to the Visual System (M_{Rep}, = 0.81, M_{Vis} = 0.77, $p = .04$). The parameter estimates of this model can be found in Table 13.

Table 13. Parameter estimates for Accuracy in Experiment 8. Significance levels: * p < .05, ** p < .01, *** p < .001

Effects	Estimate	SE	z
Intercept	2.43	0.16	14.55 ***
System (Reportative vs. Visual)	-0.62	0.30	-2.05 *
Verb vs Morpheme	-2.01	0.44	-4.50 ***
Adverb vs Morpheme	0.11	0.41	0.27

4.2.2.2. Comparison across linguistic and non-linguistic forms

So as to compare how accurately participants learned each of the linguistic forms against compared to the non-linguistic symbol (the red frame), a model similar to the one in the previous analysis was used. Participants' accuracy was included as the dependent variable along with the main effects and interaction of the type of evidential System (Visual, Reportative) and Form (verb, adverb, morpheme, frame) as the fixed predictors. Random intercepts for Participants and Items were also included. System effect was assessed by one planned comparison (contrast coding: -.50,.50). For the Form predictor, there were three planned comparisons, comparing accuracy for verbs against accuracy for frame (contrast coding: .50, 0, 0, -50), accuracy for adverbs against accuracy for frame (contrast coding: 0, .50, 0, -50) and lastly, accuracy for morphemes against accuracy for frame (contrast coding: .0, 0, .50, -50).

Including both fixed predictors in the model significantly improved the model fit based on a chi- square test of the change in -2 restricted log likelihood (χ^2 = 30.30, p= .02). There was no significant improvement when adding the interaction between the two predictors (χ^2 = 2.32, p= .50). Participants' accuracy rates were higher in the verb condition compared to the frame condition (M_{Verb}, = 0.90, M_{Frame} = 0.85, , p= .001). There was no difference in the accuracy rates between the adverb and frame conditions (M_{Adverb}, = 0.78, M_{Frame}= 0.85, p= .28). Lastly, there is a significant difference between the participants' accuracy rates for morpheme and frame conditions: there is a significantly lower accuracy rate for the morpheme condition compared to the accuracy rates for the frame condition ($M_{Morpheme}$, = 0.69, M_{Frame}= 0.85, p< .0001). Turning to the effect of System, participants performed better in the Reportative compared to the Visual System (M_{Rep}, = 0.83, M_{Vis} = 0.79, p=.01). The parameter estimates of the model can be found in Table 14.

Table 14. Parameter estimates for Accuracy across linguistic and non-linguistic forms. Significance levels: * p < .05, ** p < .01, *** p < .001

Effects	Estimate	SE	z
Intercept	2.67	0.15	17.12 ***
System (Reportative vs. Visual)	-0.65	0.27	-2.36 *
Verb vs Frame	-1.64	0.50	-3.23 **
Adverb vs Frame	2.31	0.47	4.90 ***
Morpheme vs Frame	0.49	0.46	-1.07

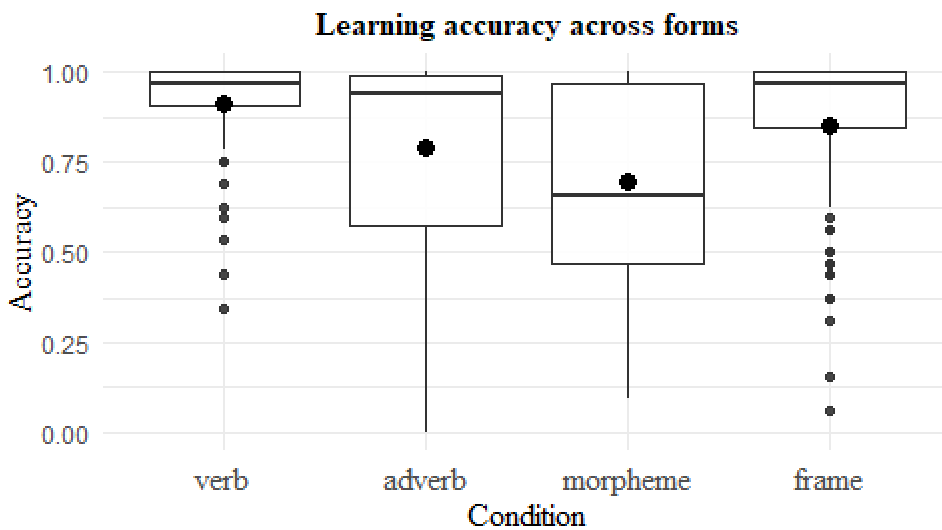

Figure 14. Accuracy score distribution, median (horizontal bar) and mean (dot) in Experiment 8.

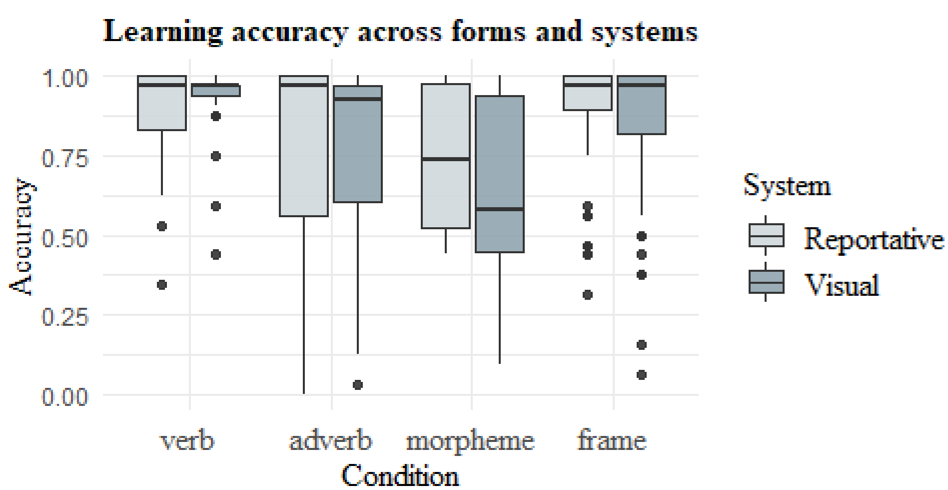

Figure 15. Accuracy score distribution and median (horizontal bar) across forms and systems in Experiment 8.

As before, participants were asked what they thought that the meaning of each novel form was and when it was used. Out of the 31 participants that were exposed to the Reportative access getting marked through a novel verb, 18 participants connected the novel form with a mentalistic notion (e.g. *think*) while 9 participants reported that the form meant that the speaker "*was told*" about what happened or that she "*heard about the event*". Additionally, 2 participants connected the novel form with "*know*" while one reported that the verb was used when the speaker "*guessed what happened*". 32 out of the 34 participants that were exposed to the verb form marking Visual access to the event reported connected the verb form with the meaning of the speaker seeing the event.

For the case of the adverb form, out of the 40 participants exposed to the Reportative evidential system, 11 participants reported that the adverbial expression meant that the speaker "*was told*" about the event (2 of which specifically pointed to meanings of "*supposedly*" and "*allegedly*". 11 participants connected it with "*secretly*" or "*not being able to see*" and 7 participants with the meaning of "*apparently*". For the Visual evidential system, 20 out of 37 participants thought that the adverb meant "*visibly*" or "*saw*" while 1 participant connected the adverbial expression with the meaning of "*certainly*". The remaining participants for each system provided irrelevant interpretations.

For the morpheme, when it marked Reportative access to the event, out of the 36 participants, 8 participants connected it with meaning that the speaker "*was told about what happened*" while 6 participants connected it with the speaker not being able to see and the event "*being hidden*". Out of the 34 participants for which the morpheme marked

Visual access, 12 participants reported that that the morpheme meant that the speaker "*saw the event*". For both Systems, the remaining participants provided meanings associations that were not relevant.

Turning to the non-linguistic symbol, the red frame, out of the 40 participants that were exposed to the frame marking Reportative access to information, 15 participants specifically associated the frame with "*the speaker being told*" about the event. Another 15 participants reported that the frame was used when the speaker "*could not see or had her eyes covered*" while 2 participants specifically connected the frame with the meaning of "*whisper"*. The rest 8 participants provided irrelevant responses. Lastly, out of the 40 participants that were exposed to the red frame marking speaker's Visual access to the event, 31 participants associated the red frame with meanings related to visual perception such "*see*", "*witness*", "*watch*" while 5 participants erroneously connected the frame with the opposite type of access (i.e. Reportative access). The rest provided irrelevant meaning interpretations.

4.2.2.3. Comparison across experiments

In an effort to examine whether the clue provided to participants in relation to the desired target meaning had any effect on their performance, we compare their learning accuracy for each condition across the two experiments. In more detail, for each condition separately, we used a mixed-effects logistic regression model that included participants' accuracy as the dependent variable along with the main effects of the type of evidential System (Visual, Reportative) and the Experiment (original instructions vs modified instructions) as the fixed predictors. Random intercepts for Participants and Items were also included. For the verb condition, neither System (Reportative, Visual) (χ^2 = 2.34, p=.12) nor Experiment (Original Instructions, Modified Instructions) (χ^2 = 3.00, p= .08) significantly improved model fit. These results suggest that participants did not perform better for neither of the two Systems (M_{Rep} = 0.92, M_{Vis} = 0.91) and their learning accuracy for the verb form did not differ across the two Experiments, ($M_{OrigInstr}$ = 0.93, $M_{ModInstr}$ = 0.90).

For the adverb condition, only System significantly improved model fit (χ^2 = 6.99, p= .0081) but not Experiment type (χ^2 = 0.19, p=.66). This pattern indicates that participants performed better for the Reportative System (M_{Rep} = 0.83, M_{Vis} = 0.76) across both experiments but there was no difference in their learning accuracy for the target adverb form ($M_{OrigInstr}$ = 0.81, $M_{ModInstr}$ = 0.78). Additionally, we also see a similar pattern for the morpheme condition: the System factors improves model fit (χ^2 = 7.25, p= .007) but not the Experiment type (χ^2 = 3.28, p=.07). This pattern signifies a significant difference in participants' learning rates for the two evidentiality systems, with higher rates for Reportative System (M_{Rep} = 0.71, M_{Vis} = 0.60) but no different for participants' learning accuracy of the form across the two experiment types ($M_{OrigInstr}$ = 0.62, $M_{ModInstr}$ = 0.70).

Lastly, for the Frame condition, the same pattern repeats: results show a significant improvement of model fit for System (Reportative vs Visual; χ^2 = 7.95, p= .004) but not for Experiment type (Original Instructions vs Modified Instructions, χ^2 = 1.87, p= .17). Such a pattern is interpreted such that across the two experiments, participants performed better for the Reportative System over the Visual (M_{Rep} = 0.88 M_{Vis} = 0.75) but their learning accuracy for the form remained roughly similar across the two experiments ($M_{OrigInstr}$ = 0.78, $M_{ModInstr}$ = 0.85).

4.3. Discussion

In Experiment 8, we used the same stimuli and procedure as in Experiment 7 with the difference that participants were provided with an explicit clue about the evidential nature of the target forms (both linguistic and non-linguistic). Experiment 8 revealed a learnability pattern strongly resembling that of Experiment 7: even when participants are provided with this disambiguation cue, it is easier for them to learn to mark evidential meanings through novel verbs compared to when these meanings are encoded through a verb-final, novel morpheme. The syntactic anchor of sentential complementation continues to create a strong form-meaning association. On the other hand, the association between the novel morpheme and the target meanings remains opaque as signified by the lower learning accuracy rates. Additionally, the additional cue did not benefit participants in how easily they learned to mark information source meanings through adverbs and morphemes. When compared to learning to mark evidential meaning through a non-linguistic symbol, the pattern observed is similar to the previous experiment: the syntactic properties of the verb create a learning advantage for learners while the structural opaqueness of the morpheme continues to have a taxing effect on participants' ability to make the target form-meaning association. Additionally, this tendency to mark reportative information sources over visual perception persists.

Overall, these findings strengthen the argument that the transparency of the mapping between forms and meanings and specifically, the syntactic properties of the utterance can influence the acquisition of the target associations. The patterns observed in these results cannot be explained by conceptual factors alone as the information source concepts remain constant across all different forms. Lastly, this learnability advantage of Reportative over Visual evidential systems confirms the pragmatic bias identified in previous findings: even when the difficulty of identifying the abstract category of evidentiality as a candidate meaning category, learners still encode certain kinds of information sources more easily compared to others, prioritizing marking a source that is pragmatically and communicative noteworthy.

4.4. General Discussion

Evidential meanings can be mapped onto different forms, apart from morphological elements. Cross-linguistically, there is a variety of lexical elements (verbs and adverbs) that can be used to express speaker's source of information and degree of certainty (Aikhenvald, 2007; 2013 Cornillie, 2007; Diewald & Smirnova, 2010; Aikhenvald & Storch, Aikhenvald, 2014; Boye, 2018; Skribnik & Kehayon, 2018; Pan, 2018). For English, perception verbs develop early as a means of connecting sensory experience with knowledge (Pillow, 1989; Pratt & Bryant, 1990; Clément, Koenig, & Harris, 2004; Ozturk & Papafragou, 2016). Specifically, the meanings of verbs that denote visual perception often get extended to refer to other cognitive processes in ways that are not true of other senses (e.g., in English, *see* can mean 'realize'; San Roque et al., 2018; Sweetser, 1990; Squartini, 2018). Additionally, adverbial expressions can also carry evidential undertones indicating speaker's information source, (e.g. *clearly*, *obviously*) with specific adverbial cases used to explicitly state that the proposition of the speaker's utterance is hearsay (e.g. *allegedly, reportedly*; Urmson, 1963; Palmer, 1986; Ifantidou-Trouki, 1993; Rocci, 2017).

Developmentally, the acquisition of grammaticalized evidentiality follows a protracted trajectory (Aksu-Koc & Alici, 2000; Unal & Papafragou, 2016; Fitneva, 2018) with children facing issues with connecting information source meanings to the target linguistic forms up to the age of six across different languages (Choi, 1995; Matsui, Yamamoto, & McCagg, 2006; Papafragou et al., 2007 ; de Villiers, Garfield, Speas, & Roeper, 2007; de Villiers et al., 2009; Turkish: Ozturk & Papafragou, 2015). There are two competing proposals about the potential factors that influence the acquisition of evidentials: under one proposal, these issues arise from conceptual difficulties due to the fact that evidentials carry epistemic overtones and express abstract concepts that cannot be easily extracted from observable events (Aksu & Slobin, 1986; Gopnik and Graf, 1988; Bartsch & Wellman, 1995; Gopnik & Meltzoff, 1997; Wiemer, 2018).On a second proposal, similarly to mental state verbs, such difficulties derive the fact that mentalistic and abstract concepts are harder to be extracted from a situational context and mapped onto the target linguistic forms (Gillette, Gleitman, Gleitman & Lederer, 1999; Snedeker & Gleitman, 2004; Gleitman, Cassidy, Nappa, Papafragou & Trueswell, 2005; Matsui et. al., 2006; Ozturk & Papafragou, 2015).

Therefore, this chapter set out to explore whether or how learning to map evidential meanings across novel linguistic forms that are meant to express evidential meanings influence the learnability patterns observed and what type of factors, if any, would influence the form-to-meaning association. So as to assess the unique effects of the mapping factors involved, we kept the conceptual demands constant: participants had to learn to express through different linguistic forms (verb, adverb or morpheme), *the same* evidential meaning (either a direct or indirect information source). Our findings indicate that there is indeed a learnability difference depending on the linguistic form that the target evidential meaning is mapped onto: participants learned more easily to map evidential meanings onto novel verbs compared to when they learn to map these meanings onto morphemes. This pattern persisted even when participants were provided with an additional cue pointing to the form-meaning association.

These consistent patterns indicate that the form on which the meanings are mapped onto in combination with the linguistic constraints that this form brings influence the learnability of the form-meaning association. More specifically, the syntactic anchor of sentential complementation that is present for the novel verb is guiding learners to the target meaning associations. However, for the case of the verb-final morpheme and the adverb, the opaqueness of the correspondence between the form and the evidential meaning does not allow participants to successfully encode and learn the target form-meaning association. Additionally, compared to learning to associate the target evidential meanings with a non-linguistic symbol (hence no linguistic constraint present), the syntactic constraints present for the novel verb actually improve the learnability of the target form-meaning association. On the other hand, no linguistic support (as in the case of the novel morpheme) has a detrimental effect on the learnability of this association.

Another interesting finding is the fact that there is also a pragmatic factor that still influences participants' learning patterns. Participants learned to mark reportative information access meanings more easily across all different forms (both linguistic and non-linguistic) than visual meanings. This is consistent with our previous findings that show that there is this pragmatic bias that motivates learners to mark sources that are potentially unreliable due to a communicative drive to be informative (Grice, 1979; Papafragou et al., 2007; Matsui & Fitneva, 2009; Aikhenvald, 2018; Wiemer, 2018).

Consistent with research in relation to mental state meanings, the syntactic structure that is associated with different linguistic forms in combination with pragmatic constraints can assist with the extraction of commonalities across evidential meanings (Gleitman, 1990, Gleitman, et. al., 2005; Papafragou, Cassidy and Gleitman, 2006; Hacquard and Lidz, 2018). Importantly, our findings show that these constraints can actually benefit or even negatively affect the learning of form-to-meaning mappings.

Chapter 5

GENERAL SUMMARY AND CONCLUSIONS

It is often assumed that cross-linguistically more frequent distinctions are easier to learn than less frequent ones (Jacobson, 1971; Rosch, 1972; Clark, 1976; Pinker, 1984; Slobin, 1985; Bowerman, 1993). In an effort to explore the different learning biases that influence learning and motivate recurring distinctions across languages, Artificial Language Learning studies expose participants to different versions of a target miniature language that differ minimally from each other in terms of a grammatical or lexical feature (see Folia, Uddén, de Vries, Forkstam, & Petersson, 2010 for a review). Such studies provide evidence that show that cross-linguistically frequent patterns are indeed easier to be learned than less common ones: this seems to be true for syntactic (Christiansen, 2000; Newport & Aslin, 2004; Hudson Kam & Newport, 2005, 2009; Thompson & Newport, 2007; Wonnacott, Newport, & Tanenhaus, 2008; Tily, Frank, & Jaeger, 2011; Culbertson, 2012; Culbertson & Smolensky, 2012; Culbertson, Smolensky, & Legendre, 2012), phonological (Seidl & Buckley, 2005; Wilson, 2006; Finley & Badecker, 2009) and phonological distinctions (Merkx, Rastle, & Davis, 2011; Fedzechkina, Jaeger, & Newport, 2012). However, very few studies have provided empirical evidence on semantic distinctions (Xu, Dowman & Griffiths, 2013; Carstensen, Xu, Smith & Regier 2015; Maldonado & Culbertson, 2019).

In this work, we provide the first systematic study of the potential factors that contribute to the learnability and the cross-linguistic prevalence of certain semantic distinctions over others. Within the semantic domain, we focus on *evidentiality* since it offers an excellent test case for learnability patterns due to the cross-linguistic frequency of certain evidential systems. Additionally, the fact that evidentiality is not a grammaticalized phenomenon in English allows for such distinctions to be taught to English-speaking adults within an Artificial Language Learning paradigm without native language interference.

5.1. Summary of the findings and theoretical connections

Under Gentner and Bowerman's Typological Prevalence Hypothesis, the distinctions that are cross-linguistically more prevalent are easier to learn due to the fact that they reflect more 'natural' conceptual categories. In Chapter 2, we put tested this hypothesis using evidential distinctions as our test case. Across languages that grammatically mark only one type of information, evidential systems that involve only Reportative morphemes are the most widespread ones. Systems that use an indirect morpheme to mark inference or reports are less frequent (Papafragou et al., 2007; deHaan, 2013a; Aikhenvald, 2004, 2018; Ünal & Papafragou, 2018;). Evidential systems that only have Visual morphemes are rare (Aikhenvald, 2018). Therefore, this crosslinguistic frequency pattern provided an ideal test base to compare the learnability of these distinctions through an Artificial Language Learning Paradigm.

As the TPH predicted, when exposed to these miniature evidential morphological systems, adult speakers of English whose language does not encode evidentiality grammatically learned the typologically most prevalent system (marking indirect, reportative information) better compared to less-attested systems (Experiments 1-2). Similar patterns were observed when non-linguistic symbols were used to encode evidential distinctions (Experiment 3). However, the ubiquity and cross-linguistic prevalence of visual perception as a source of knowledge (Viberg, 1984; Pillow, 1989; Pratt & Bryant, 1990; San Roque et al., 2015; Ozturk & Papafragou, 2016) contradicts the proposal that TPH sets forward that the most easily learned distinctions derives from the naturalness of the underlying concepts. Instead, we propose that these facts reflect either major, internally coherent categories of meaning distinctions for human cognizers, or pragmatic pressures on which types of linguistic meanings are informative to encode (or possibly a combination of these two factors). Depending on the relative contribution of such competing factors, frequently expressed and easy-to-acquire linguistic meanings can be a window onto conceptual categories but can alternatively or additionally reveal the role of pragmatic mechanisms within cognitive architecture.

In Chapter 3, the aim was to adjudicate between a pragmatic bias and category partition issues as an explanation for the learnability of evidential distinctions. In an effort to eliminate design complications in relation to the difficulty of differentiating visual from inferential access to information in the specific scenarios involved, we exposed participants to miniature evidential morphological systems that used a novel morpheme to mark either only indirect access to information (reportative access) or direct access to information (visual access). Adult English speakers consistently showed higher learning rates for systems that marked indirect compared to direct information sources (Experiment 1). This pattern persisted even when participants were given specific cues to the target evidential meanings (Experiment 2) and partly extended to cases where participants had to encode evidential meanings through visual, not linguistic, means (Experiment 3).

These findings show that certain semantic distinctions are easier to learn (and hence more prevalent across different languages) not due to the naturalness of the underlying meanings being expressed but because of their communicative contribution. There is a pragmatic motivation for participants to mark the information sources that are potential unreliable (Koenig, & Harris, 2005; Jaswal, & Neely, 2006; Mascaro & Sperber, 2009; Koenig, 2012; Jaswal, 2010; Fusaro, Corriveau, & Harris, 2011; Harris, 2012;Arslan, 2020; Degen et al., 2019) in an effort to be maximally informative (Grice, 1975; Sperber & Wilson, 1986).

In our final chapter, Chapter 4, we provided evidence as to how factors related to the mapping process between the linguistic expressions and the target meanings can influence the learnability of semantic distinctions. Using an artificial language learning paradigm, we compare adult learners' acquisition of a single evidential meaning expressed by different linguistic forms (a novel verb/morpheme/adverb). The goal was to see whether mapping the same concept onto different forms yields different learning outcomes. This would eliminate concerns about the differential complexity of the concepts involved (Aksu & Slobin, 1986; Gopnik and Graf, 1988; Bartsch & Wellman, 1995; Gopnik & Meltzoff, 1997; Wiemer, 2018) and instead highlight the unique effects of the mapping factors involved in forming the target form-meaning association. In our findings, it seems that syntactic and pragmatic constraints have a beneficial effect on learning the target form-meaning association, even in the presence of an additional cue that specifically guided learners towards the target meanings. Such findings provide novel evidence in support of the claim that what makes certain meanings easy or hard to learn, regardless of their conceptual presuppositions, often lies in the transparency of the correspondence between those meanings and the linguistic forms that express them.

Overall, we take the outcome of this dissertation work as a promising framework for understanding how foundational pragmatic factors shape the acquisition of semantics in ways that complement the more heavily studied conceptual factors underpinning language structure and acquisition. How to disentangle the contributions of conceptual from pragmatic (and other mapping) factors to the acquisition process is an exciting direction for future work.

REFERENCES

Aikhenvald, A. (2004). *Evidentiality*. Oxford University Press.

Aikhenvald, A. (2014). The grammar of knowledge: a cross-linguistic view of evidentials, and the expression of information source. In A. Aikhenvald and R.M.W. Dixon (Eds.), *The grammar of knowledge: A cross-linguistic typology* (pp. 1–51). Oxford: Oxford University Press.

Aikhenvald, A.(2018). Evidentiality: The Framework. In A. Aikhenvald (Ed.), *The Oxford Handbook of Evidentiality* (pp.1-43). Oxford: Oxford University Press.

Aikhenvald, A., & Dixon, R. (Eds.). (2003). *Studies in evidentiality*. Amsterdam: Benjamin.

Aikhenvald, A., & Dixon, R. (Eds.). (2014). *The grammar of knowledge: A cross-linguistic typology*. Oxford: Oxford University Press.

Aksu-Koç, A. (1988). *The acquisition of aspect and modality: The case of past reference in Turkish*. Cambridge, UK: Cambridge University Press.

Aksu-Koç, A., & Alici, D. (2000). Understanding sources of beliefs and marking of uncertainty. In E. Clark (Ed.), *Proceedings of the 30th Annual Child Language Research Forum* (pp. 123–130). Stanford, CA: CSLI.

Aksu-Koç, A., Ögel-Balaban, H., & Alp, E. I. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development*, *125*,13–28. https://doi.org/10.1002/cd.247.

AnderBois, S.(2014). On the exceptional status of reportative evidentials. *Proceedings of SALT*, *24,* 234-54.

Arslan, S. (2020). When the owner of information is unsure: Epistemic uncertainty influences evidentiality processing in Turkish. *Lingua* (247).

Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511801686

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390 – 412.

Bannard, C., Rosner, M., & Matthews, D. (2017). What's worth talking about? Information theory reveals how children balance informativeness and ease of production. *Psychological Science, 28*(7), 954 – 966. doi:10.1177/0956797617699848

Bartsch, R., & Wellman, H. (1995). *Children talk about the mind*. Oxford, England: Oxford University Press

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bedny, M., Koster-Hale, J., Elli, G., Yazzolino, L., & Saxe, R. (2019). There's more to "sparkle" than meets the eye: knowledge of vision and light verbs among congenitally blind and sighted individuals. *Cognition*, *189*, 105–115. https://doi.org/10.1016/j.cognition.2019.03.017

Bowerman, M. (1993). Typological perspectives on language acquisition: Do crosslinguistic patterns predict development? In E. V. Clark (Ed.), *Proceedings of the 25th Annual Child Language Research Forum* (pp. 7–15). Stanford: Center for the Study of Language and Information.

Brosseau-Liard, P. E., & Birch, S. A. J. (2011). Epistemic states and traits: Preschoolers appreciate the differentialinformativeness of situation-specific and person-specific cues to knowledge. *Child Development, 82*, 1788–1796.

Boye, K. (2018). Evidentiality: the notion and the term. In A. Aikhenvald (Ed.), *The Oxford handbook of evidentiality* (pp. 261–272). Oxford, UK: Oxford University Press.

Carr, J. W., Smith, K., Culbertson, J., & Kirby, S. (2020). Simplicity and informativeness in semantic category systems. *Cognition*, *202*, 104289–104289. https://doi.org/10.1016/j.cognition.2020.104289

Carstensen, A., Xu, J., Smith, C., and Regier, T. (2015). Language evolution in the lab tends toward informative communication. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX. Cognitive Science Society.

Chafe, W., & Nichols, J. (Eds.). (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.

Christiansen, M.H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In J.L. Dessalles and L. Ghadakpour (Eds.) *The Evolution of Language: 3rd International Conference* (pp. 45-48). Paris: France.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.

Clark, E. (1976). Universal categories: On the semantics of classifiers and children’s early word meanings. In Juilland, A. (Ed), *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday* (Vol. 3, pp. 449–462). Saratoga, CA: Anna Libri.

Clark, H. (1973). Space, time semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press.

Cornillie, B. (2007). *Epistemic modality and evidentiality in Spanish (semi-) auxiliaries: a cognitive-functional approach*. Berlin: Mouton de Gryer.

Cornillie, B., Arrese, M., J. & Wiemer, B. (2015). Evidentiality and the semantics-pragmatics interface. *Belgian Journal of Linguistics*, *29*, 1-17. doi 10.1075/bjl.29.001int

Culbertson, J. (2012). Typological Universals as Reflections of Biased Learning: Evidence from Artificial Language Learning. *Linguistics and Language Compass*, *6*(5), 310–329. https://doi.org/10.1002/lnc3.338

Culbertson, J., & Smolensky, P. (2012). A Bayesian Model of Biases in Artificial Language Learning: The Case of a Word-Order Universal. *Cognitive Science*, *36*(8), 1468–1498. https://doi.org/10.1111/j.1551-6709.2012.01264.x

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329. https://doi.org/10.1016/j.cognition.2011.10.017

Dancy, J. (1985). An introduction to contemporary epistemology. Oxford: Blackwell. Delancey, S. (2002). The mirative and evidentiality. *Journal of Pragmatics*, *33*, 369–382.

de Haan, F. (2001). The place of inference within the evidential system. *International Journal of American Linguistics*, *67*, 193-219.

de Haan, F. (2013). Coding of evidentiality. In M. Dryer and M. Haspelmath, (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

de Villiers, J.G. , Garfield, J., Gernet Girard, H., Roeper, T. & Speas, P. (2009). Evidentials in Tibetan: Acquisition, semantics and cognitive development. *New Directions for Adolescent and Child Development* (Special Issue on evidentiality), *125*, 29-48.

Degen, J., Trotzke, A., Scontras, G., Wittenberg, E., & Goodman, N. D. (2019). Definitely, maybe: A new experimental paradigm for investigating the pragmatics of evidential devices across languages. *Journal of Pragmatics*, *140*, 33–48. https://doi.org/10.1016/j.pragma.2018.11.015

Delancey, S. (2002). The mirative and evidentiality. *Journal of Pragmatics*, *33,* 369–382.

Diewald, G. & Sminorva, E. (2010). Evidentiality in European languages. The lexical grammatical distinction. In G. Diewald & E. Smirnova (Eds.), *Linguistic realization of evidentiality in European languages* (pp. 1-14). Berlin: Mouton de Gryer.

Faller, M. (2012). Evidential scalar implicatures. *Linguistics and Philosophy*, *35*, 285–312. https://doi.org/10. 1007/s10988-012-9119-8.

Faller, M. (2014). Reportative evidentials and modal subordination. *Lingua*, *186*,55–67. https://doi.org/10. 1016/j.lingua.2014.10.004.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897–17902. https://doi.org/10.1073/pnas.1215776109

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2016). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science*, *41*(2), 416–446. https://doi.org/10.1111/cogs.12346

Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, *61*(3), 423–437. https://doi.org/10.1016/j.jml.2009.05.002

Fitneva, S. (2018). The acquisition of evidentiality. In A. Aikhenvald (Ed.), *The*

Oxford handbook of evidentiality (pp. 185–201). Oxford, UK: Oxford University Press.

Fitneva, S. A. (2008). The role of evidentiality in Bulgarian children's reliability judgments. *Journal of Child Language*, *35*, 845–868. https://doi.org/10.1017/S0305000908008799.

Fitneva, S. A., Lam, N. H. L., & Dunfield, K. A. (2013). The development of children's information gathering: To look or to ask? *Developmental Psychology*, *49*, 533–542.

Folia, V., Uddén, J., de Vries, M. H., Forkstam, C., & Petersson, K. M. (2010). Artificial language learning in adults and children. *Language Learning*, *60*(s2), 188–220.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96. doi:10.1016/j.cogpsych.2014.08.002

Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology*, *110*(4), 561–574. https://doi.org/10.1016/j.jecp.2011.06.008

Gärdenfors, P. 2014. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge, MA: MIT Press.

Garrett, E. (2000). *Evidentiality and assertion in Tibetan*. UCLA doctoral diss.

Gentner, D. and Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, and S. Özçalıskan (Eds), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 465–480). Hillsdale, NJ: Lawrence Erlbaum

Associates.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S.T., & Conway, B.R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785-10790.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135–176.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, *59*, 1366–1371.

Gleitman, L. R. (1990). The structural sources of verb meanings. Language Acquisition, 1, 3–55.

Gleitman, L., Cassidy, K., Nappa, R., Papafragou, A. & Trueswell, J. (2005). Hard words. *Language Learning and Development*, 1(1), 23– 64.

Greenfield, P. M. (1979). Informativeness, presupposition, and semantic choice in single-word utterances. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 159–166). New York, NY: Academic Press.

Greenfield, P. M., & Smith, J. H. (1976). *The structure of communication in early language development.* New York, NY: Academic Press.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grigoroglou, M., Johanson, M., & Papafragou, A. (2019). Pragmatics and spatial language: The acquisition of *front* and *back*. *Developmental Psychology*, *55*, 729-744.

Hacquard, V., & Lidz, J. (2019). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind and Language*, *34*(1), 73–96.

https://doi.org/10.1111/mila.12192.

Hannigan, S. L., & Reinitz, M. T. (2001). A Demonstration and Comparison of Two Types of Inference-Based Memory Errors. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(4), 931–940. https://doi.org/10.1037/0278-7393.27.4.931

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*, 139–151. https://doi.org/10.1006/anbe.2000.1518.

Harris, P. L. (2012). *Trusting what you are told*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Haspelmath, M. (1997). *Indefinite Pronouns*. London: Clarendon.

Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151–195.

Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66.

Ifantidou-Trouki, E. (1993). Sentential adverbs and relevance. *Lingua*, *90*(1), 69–90. https://doi.org/10.1016/0024-3841(93)90061-Z

Ifantidou, E. (2001). *Evidentials and relevance*. Amsterdam: Benjamin.

Izvorski, R. (1998). The present perfect as an epistemic modal. In A. Lawson & E. Cho (Eds.), *Proceedings of SALT 7*. Ithaca, NY: CLC Publications.

Jakobson, R. (1971). *Studies on child language and aphasia*. The Hague/Paris: Mouton.

Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology*, *61*, 248–

272.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*, 757–758.

Johanson, L., & Utas, B. (Eds.). (2000). *Evidentials: Turkic, Iranian and neighbouring languages*. Berlin: de Gruyter.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3– 28. https://doi.org/10.1037/0033-2909.114.1.3.

Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language*, *72*,533–78.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109-128.

Koenig, M. A. (2012). Beyond Semantic Accuracy: Preschoolers Evaluate a Speaker's Reasons. *Child Development*, *83*(3), 1051–1063. https://doi.org/10.1111/j.1467-8624.2012.01742.x.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mis- trust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277.

Koring, L., & De Mulder, H. (2015). Understanding different sources of information: the acquisition of evidentiality. *Journal of Child Language*, *42*, 947–968. https://doi.org/10.1017/S030500091400052X.

Koster-Hale, J., Bedny, M., & Saxe, R. (2014). Thinking about seeing: Perceptual sources of knowledge are encoded in the theory of mind brain regions of sighted and blind adults. *Cognition*, *133*(1), 65–78. https://doi.org/10.1016/j.cognition.2014.04.006.

Kratzer, A. (1991). Modality. In A. von Stechow & D. Wunderlich (Eds.), *Semantics: An International Handbook of Contemporary Research* (pp. 639–650). Berlin:

de Gruyter.

Landau, B. & Gleitman, L.R. (1985). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Lee, T., & Law, A. (2000). Evidential final particles in child Cantonese. In E. Clark (Ed.), *Proceedings of the 30th Annual Child Language Research Forum* (pp. 131–138). Stanford, CA: CSLI.

Levinson, S.C., & Meira, S. (2003). 'Natural concepts' in the spatial topologial domain—adpositional meanings in crosslinguistic perspective: an exercise in semantic typology. *Language, 79*, 485–516.

Majid, A., Boster, J.S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: a study of cutting and breaking. *Cognition*, *109*, 235–50.

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, B., De Sousa, H., Cansler, B. L., Shayan, S., De Vos, C., Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemanse, M., Ozturk, O., Brown, P., Hill, C., Le Guen, O., Hirtzel, V., Van Gijn, R., Sicoli, M. A., & Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11369-11376.

Maldonado, M. & Culbertson, J. (2019). Something about "us": Learning first person pronoun systems. In Proceedings of the 41st Annual Meeting of the Cognitive Science Society. Cognitive Science Society, Austin, TX.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*(3), 367–380. https://doi.org/10.1016/j.cognition.2009.05.012.

Matsui, T. & Fitneva, S. A. (2009). Knowing how we know: evidentiality and cognitive development. In S. A. Fitneva & T. Matsui (eds), *Evidentiality: a* *window into language and cognitive development* (New Directions for Child and Adolescent Development 125) (pp. 1–11). San Francisco: Jossey-Bass.

Matsui, T., Miura, Y., & McCagg, P. (2006a). Young children's implicit and explicit understanding of speaker knowledge. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Cognitive Science Society* (pp. 1789–1794). Mahwah, NJ: Lawrence Erlbaum.

Matsui, T., Yamamoto, T., & McCagg, P. (2006b). On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development*, *21*, 158–173. https://doi.org/10.1016/j.cogdev.2005. 10.001.

Matthewson, L. (2012). When Fieldwork Meets Theory: Evidence About Evidentials. In B. Stolterfoht & S. Featherston (Eds.), *Empirical Approaches to Linguistic Theory: Studies of Meaning and Structure*, (pp. 85–114). Berlin, Germany: De Gruyter.

McCready, E. (2008). Semantic heterogeneity in evidentials. New Frontiers in Artificial Intelligence, LNCS 4914.

McCready, E., & Ogata, N. (2007). Evidentiality, modality, and probability. *Linguistics and Philosophy*, *30*(2), 147–206.

Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *Quarterly Journal of Experimental Psychology*, *64*(6), 1200–1220. https://doi.org/10.1080/17470218.2010.538211

Murray, E., S. (2011). A Hamblin semantics for evidentials. In C., S., Ito & D. Lutz (Eds.), *Proceedings from Semantics and Linguistic Theory (SALT) XIX* (2009), 324-41. Ithaca, NY: CLC Publications.

Murray, E., S. (2014). Varieties of update. *Semantics and Pragmatics*, *7,* 1-53.

Mushin, I. (2001). *Evidentiality and epistemological stance: Narrative retelling*.

Amsterdam: Benjamin.

Newport, E. and Aslin, R. (2004). Learning at a distance in statistical learning of non-adjacent dependencies. *Cognitive psychology*, *48*(2):127–162.

O'Neill, D., & Astington, J. (1990, April). Preschooolers' developing understanding of the role sensory experiences play in knowledge acquisition. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

O'Neill, D., & Chong, S. (2001). Preschool children's difficulty understanding the types of information obtained through the five senses. *Child Development*, *72*, 803–815.

Ozturk, O., & Papafragou, A. (2016). The Acquisition of Evidentiality and Source Monitoring. *Language Learning and Development*, *12*(2), 199–230. https://doi.org/10.1080/15475441.2015.1024834

Palmer, F., 1986. *Mood and modality*. Cambridge: Cambridge University Press.

Pan, C. (2018). Evidentiality in Formosan languages. In A. Aikhenvald (Ed.), *The Oxford handbook of evidentiality* (pp. 629–656). Oxford, UK: Oxford University Press.

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition*, *105*, 125–165. https://doi.org/10.1016/j.cognition.2006.09.008

Papafragou, A., Li, P., Choi, Y., & Han, C. H. (2007). Evidentiality in language and cognition. *Cognition*, *103*, 253–299. https://doi.org/10.1016/j.cognition.2006.04.001.

Pillow, B. (1989). Early understanding of perception as a source of knowledge. Journal of *Experimental Child Psychology*, *47*, 116–129.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Plungian, V. A. (2001). The place of evidentiality within the universal grammatical space. *Journal of Pragmatics, 33,* 349–357. doi: 10.1016/S0378-2166(00)00006-0

Pyers, J. E., & Senghas, A. (2009). Language promotes false- belief understanding: Evidence from learners of a new sign language. *Psychological Science*, *20*, 805–812. https://doi.org/10.1111/j.1467-9280.2009. 02377.x.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: words for snow revisited. *Plos One*, 11(4), 0151138. https://doi.org/10.1371/journal.pone.0151138

Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. M. W. O'Grady (Ed.), The handbook of language emergence (pp. 237– 263). Hoboken, NJ: Wiley-Blackwell.

Rissman, L., Woodward, A., & Goldin-Meadow, S. (2019). Occluding the face diminishes the conceptual accessibility of an animate agent. *Language, Cognition and Neuroscience*, *34*(3), 273–288. https://doi.org/10.1080/23273798.2018.1525495.

Robinson, E., Champion, H., & Mitchell, P. (1998). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology*, *35*, 535–546.

Robinson, E., Haigh, S., & Nurmsoo, E. (2008). Children's working understanding of knowledge sources: Confidence in knowledge gained from testimony. *Cognitive Development*, *23*, 105–118.

Rocci, A. (2017). *Modality in argumentation : a semantic investigation of the role of modalities in the structure of arguments with an application to italian modal expressions* (Ser. Argumentation library, v. 29). Springer.

https://doi.org/10.1007/978-94-024-1063-1

Rosch, E. H. (1972). Universals in color naming and memory. *Journal of experimental psychology*, *93*(1):10–20.

Sabbagh, M.A., & Baldwin, D.A. (2001). Learning words from knowl- edgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054–1070.

San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., Dirksmeyer, T., Enfield, NJ., Floyd, S., Hammond, J., Rossi, G., Tufvesson, S., vanPutten, S. & Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics*, *26*, 31–60. https://doi.org/10.1515/cog-2014-0089.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. Language Learning and Development, 1, 289–316.

Skribnik E. & Kehayov, P. (2018). Evidential in Uralic languages. In A. Aikhenvald (Ed.), *The Oxford handbook of evidentiality* (pp. 525–553). Oxford, UK: Oxford University Press.

Snedeker, J. & Gleitman, L. (2004). Why it is hard to label our concepts. In S. Waxman & G. Hall (Eds.), *Weaving a lexicon*. Cambridge, MA: The MIT Press.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues* (pp. 1157–1256). Hillsdale, NJ: Lawrence Erlbaum Associates.

Speas, M. (2004). Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua*, *114*, 255–276.

Speas, M. (2018). Evidentiality and formal semantic theories. In A. Aikhenvald (Ed.) *The Oxford Handbook of Evidentiality*, (pp. 286-311) Oxford: Oxford University

Press.

Sperber, D., Cara, F., & V. Girotto (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31-95.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010), Epistemic Vigilance. *Mind & Language*, *25*, 359-393. https://doi-org.udel.idm.oclc.org/10.1111/j.1468-0017.2010.01394.x.

Squartini, M. (2018). Extragrammatical expression of information source. In A. Aikhenvald (Ed.), *The Oxford Handbook of Evidentiality* (pp. 273-285), Oxford: Oxford University Press.

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*(3), 409–415. https://doi.org/10.1016/j.cognition.2011.04.007.

Tabullo, Á., Arismendi, M., Wainselboim, A., Primero, G., Vernis, S., Segura, E., … Yorio, A. (2012). On the learnability of frequent and infrequent word orders: An artificial language learning study. *Quarterly Journal of Experimental Psychology*, *65*(9), 1848–1863. https://doi.org/10.1080/17470218.2012.677848

Thompson, S, and E.L. Newport. (2007). Statistical learning of syntax: The role of transitional probability. Language *Learning and Development*, *3*, 1-42.

Tily, H., Frank, M., and Jaeger, F. (2011). The learnability of constructed languages reflects typological patterns. In L. Carlson, C. Hoelscher and T. Shipley (Eds), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp.1364–1369). Boston, MA: Cognitive Science Society.

Ünal, E., & Papafragou, A. (2016). Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, *89*, 179–199. https://doi.org/10.1016/j.jml.2015.12.001.

Ünal, E., & Papafragou, A. (2018). Relations between language and cognition:

Evidentiality and sources of knowledge. *Topics in Cognitive Science*, 1–21. Advance Online Publication. https://doi.org/10.1111/tops.12355.

Ünal, E., & Papafragou, A. (2019). How children identify events from visual experience. *Language Learning and Development* 15, 138-156.

Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language*, *87*, 157–176. https://doi.org/10.1016/j.jml.2015. 10.009.

Urmson, J., 1963. Parenthetical verbs. In C. Caton (Ed.), *Philosophy and ordinary language*, (220-240). Urbana, IL: University of Illinois Press

Uzundag, B. A., Taşçı, S. S., Küntay, A. C., & Aksu-Koç, A. (2018). Functions of Turkish evidentials in early child-caregiver interactions: A growth curve analysis of longitudinal data. *Journal of Child Language, 45*(4), 878-899 https://doi.org/10.1017/s0305000911000183.

Van der Auwera J., & Plungian, V. (1998). Modality's semantic map. *Linguistic Typology*, *2*, 79–124.

Van der Henst, J-B., Bujakowska K., Ciceron C., & Noveck I.A.. (2006). How to make a participant logical: The role of premise presentation in a conditional reasoning task. In D. Andler, Y. Ogawa, M. Okada & S. Watanabe (Eds.), *Reasoning and Cognition* (pp.7-17). Keio University.

Viberg, Å. (1984). The verbs of perception: A typological study. In B. Butterworth, B. Comrie and Ö. Dahl (Eds.), *Explanations for Language Universals*, (pp.123–162). Berlin: Mouton de Gruyter.

Wiemer, B. (2018). Evidentials and epistemic modality. In A. Aikhenvald (Ed.), *The Oxford Handbook of Evidentiality* (pp. 85-108), Oxford: Oxford University Press.

Willett, T. (1988). A cross-linguistic survey of the grammaticization of evidentiality.

Studies in language, *12*(1):51–97.

Wilson, C. (2006). An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982.

Winans, L., Hyams, N., Rett, J., & Kalin, L. (2014). Children's comprehension of syntactically encoded evidentiality. Paper presented at the 45th annual meeting of the North East Linguistic Society, Cambridge, MA.

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, *95*, 36–48. https://doi.org/10.1016/j.jml.2017.01.005

Woolley, J., & Bruell, M. (1996). Young children's awareness of the origins of their mental representations. *Developmental Psychology*, *32*, 335–346

Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. Proceedings of the Royal Society B: Biological Sciences, 280(1758), 20123073.

APPENDIX A

LIST OF EVENTS USED IN EXPERIMENT 1

Events			
1	The frog erased the board.	**21**	The gorilla mixed the salad.
2	The gorilla shelled the peanuts.	**22**	The giraffe grated the cheese.
3	The squirrel transferred the strawberries.	**23**	The mouse chopped the celery.
4	The gorilla prepared the s'more.	**24**	The giraffe labeled the cards.
5	The squirrel opened the drawer.	**25**	The frog drew the square.
6	The gorilla colored the star.	**26**	The squirrel gathered the flowers.
7	The frog decorated the cupcake.	**27**	The giraffe plucked the petals.
8	The frog folded the tortillas.	**28**	The mouse removed the coughdrops.
9	The frog sliced the pepper.	**29**	The mouse crumpled the papers.
10	The squirrel emptied the pouch.	**30**	The doggy replaced the eggs.
11	The squirrel sorted the crayons.	**31**	The giraffe aligned the figurines.
12	The gorilla drank the juice.	**32**	The doggy split the crackers.
13	The gorilla inflated the balloon.	**33**	The doggy peeled the banana.
14	The frog stacked the cups.	**34**	The doggy hammered the nails.
15	The frog picked the grapes.	**35**	The gorilla skinned the potato.
16	The gorilla husked the corn.	**36**	The squirrel connected the dots.
17	The squirrel poured the cereal.	**37**	The frog ate the apple.
18	The frog completed the crossword.	**38**	The gorilla strung the beads.
19	The squirrel built the steps.	**39**	The mouse solved the puzzle.
20	The squirrel sharpened the pencils.		

APPENDIX B

LIST OF EVENTS USED IN EXPERIMENTS 2 AND 3

Events					
1	The girl divided the rings.	**24**	The girl dissolved the tablet.	**47**	The girl clipped the sheets.
2	The girl squeezed the toothpaste.	**25**	The girl piled the cards.	**48**	The girl folded the towel.
3	The girl unzipped the jacket.	**26**	The girl circled the word.	**49**	The girl colored the star.
4	The girl filled the bowl.	**27**	The girl packaged the foam.	**50**	The girl cracked the peanut.
5	The girl dumped the trash.	**28**	The girl closed the book.	**51**	The girl strung the beads.
6	The girl solved the puzzle.	**29**	The girl layered the stickers.	**52**	The girl shoveled the dirt.
7	The girl undressed the doll.	**30**	The girl pasted the photos.	**53**	The girl separated the plates.
8	The girl sorted the morphemes.	**31**	The girl unfurled the umbrella.	**54**	The girl aligned the animals.
9	The girl bit the apple.	**32**	The girl flattened the putty.	**55**	The girl drew the square.
10	The girl flipped the pans.	**33**	The girl suspended the curtains.	**56**	The girl deflated the balloon.
11	The girl laid the bricks.	**34**	The girl built the tower.	**57**	The girl popped the coughdrops.
12	The girl fanned the spoons.	**35**	The girl erased the board.	**58**	The girl stapled the packet.
13	The girl untied the shoe.	**36**	The girl copied the drawing.	**59**	The girl connected the dots.
14	The girl wound the yarn.	**37**	The girl formed the snowman.	**60**	The girl emptied the pouch.
15	The girl rotated the vase.	**38**	The girl unscrewed the lightbulb.	**61**	The girl arranged the blocks.
16	The girl decorated the cookie.	**39**	The girl unrolled the shirt.	**62**	The girl removed the letter.
17	The girl chopped the celery.	**40**	The girl poured the juice.	**63**	The girl braided the string.

18	The girl lit the lamp.	**41**	The girl unfolded the napkin.	**64**	The girl gathered the flowers.
19	The girl hung the hanger.	**42**	The girl transferred the candies.	**65**	The girl stacked the cups.
20	The girl pierced the page.	**43**	The girl cleared the table.	**66**	The girl sliced the bread.
21	The girl mixed the gumballs.	**44**	The girl peeled the banana.	**67**	The girl unpacked the bookbag.
22	The girl sealed the envelope.	**45**	The girl cut the triangle.	**68**	The girl unbuttoned the sweater.
23	The girl crumpled the papers.	**46**	The girl matched the socks.	**69**	The girl split the cracker.

APPENDIX C

LIST OF EVENTS USED IN EXPERIMENTS 4-8

Events			
1	She squeezed the toothpaste.	**22**	She cracked the peanut.
2	She sorted the markers.	**23**	She strung the beads.
3	She bit the apple.	**24**	She shoveled the dirt.
4	She flipped the pans.	**25**	She separated the plates.
5	She wound the yarn.	**26**	She aligned the animals.
6	She rotated the vase.	**27**	She drew the square.
7	She chopped the celery.	**28**	She deflated the balloon.
8	She lit the lamp.	**29**	She popped the coughdrops.
9	She hung the hanger.	**30**	She stapled the packet.
10	She mixed the gumballs.	**31**	She connected the dots.
11	She circled the word.	**32**	She emptied the pouch.
12	She packaged the foam.	**33**	She arranged the blocks.
13	She closed the book.	**34**	She matched the socks.
14	She layered the stickers.	**35**	She removed the letter.
15	She pasted the photos.	**36**	She braided the string.
16	She unfurled the umbrella.	**37**	She gathered the flowers.
17	She flattened the putty.	**38**	She stacked the cups.
18	She suspended the curtains.	**39**	She sliced the bread.
19	She clipped the sheets.	**40**	She unpacked the bookbag.
20	She folded the towel.	**41**	She unbuttoned the sweater.
21	She colored the star.	**42**	She split the cracker.

APPENDIX D

IRB/HUMAN SUBJECTS APPROVAL

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	July 6, 2017
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-17] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	July 6, 2017
EXPIRATION DATE:	July 24, 2018
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

- 1 -

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	June 28, 2018
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-18] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	June 28, 2018
EXPIRATION DATE:	July 24, 2019
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

Institutional Review Board
210H Hullihen Hall
Newark, DE 19716
Phone: 302-831-2137
Fax: 302-831-2828

DATE:	August 5, 2019
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-20] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	August 5, 2019
EXPIRATION DATE:	July 24, 2020
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your Continuing Review/Progress Report submission to the University of Delaware Institutional Review Board (UD IRB). The UD IRB has reviewed and APPROVED the proposed research and submitted documents via Expedited Review in compliance with the pertinent federal regulations.

As the Principal Investigator for this study, you are responsible for and agree that:

- All research must be conducted in accordance with the protocol and all other study forms as approved in this submission. Any revisions to the approved study procedures or documents must be reviewed and approved by the IRB prior to their implementation. Please use the UD amendment form to request the review of any changes to approved study procedures or documents.
- Informed consent is a process that must allow prospective participants sufficient opportunity to discuss and consider whether to participate. IRB-approved and stamped consent documents must be used when enrolling participants and a written copy shall be given to the person signing the informed consent form.
- Unanticipated problems, serious adverse events involving risk to participants, and all non-compliance issues must be reported to this office in a timely fashion according with the UD requirements for reportable events. All sponsor reporting requirements must also be followed.

Oversight of this study by the UD IRB REQUIRES the submission of a CONTINUING REVIEW seeking the renewal of this IRB approval, which will expire on July 24, 2020. A continuing review/progress report form and up-to-date copies of the protocol form and all other approved study materials must be submitted to the UD IRB at least 45 days prior to the expiration date to allow for the required IRB review of that report.

If you have any questions, please contact the UD IRB Office at (302) 831-2137 or via email at hsrb-research@udel.edu. Please include the study title and reference number in all correspondence with this office.

Institutional Review Board
210H Hullihen Hall
Newark, DE 19716
Phone: 302-831-2137
Fax: 302-831-2828

DATE:	August 27, 2020
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-22] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	August 27, 2020
EXPIRATION DATE:	July 24, 2021
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your Continuing Review/Progress Report submission to the University of Delaware Institutional Review Board (UD IRB). The UD IRB has reviewed and APPROVED the proposed research and submitted documents via Expedited Review in compliance with the pertinent federal regulations.

Please continue to reference https://research.udel.edu/coronavirus for the most up-to-date recommendations regarding in-person research interaction with subjects during the COVID-19 national emergency.

As the Principal Investigator for this study, you are responsible for and agree that:

- All research must be conducted in accordance with the protocol and all other study forms as approved in this submission. Any revisions to the approved study procedures or documents must be reviewed and approved by the IRB prior to their implementation. Please use the UD amendment form to request the review of any changes to approved study procedures or documents.
- Informed consent is a process that must allow prospective participants sufficient opportunity to discuss and consider whether to participate. IRB-approved and stamped consent documents must be used when enrolling participants and a written copy shall be given to the person signing the informed consent form.
- Unanticipated problems, serious adverse events involving risk to participants, and all non-compliance issues must be reported to this office in a timely fashion according with the UD requirements for reportable events. All sponsor reporting requirements must also be followed.

Oversight of this study by the UD IRB REQUIRES the submission of a CONTINUING REVIEW seeking the renewal of this IRB approval, which will expire on July 24, 2021. A continuing review/progress report form and up-to-date copies of the protocol form and all other approved study materials must be submitted to the UD IRB at least 45 days prior to the expiration date to allow for the required IRB review of that report.

If you have any questions, please contact the UD IRB Office at (302) 831-2137 or via email at hsrb-research@udel.edu. Please include the study title and reference number in all correspondence with this office.

Institutional Review Board
210H Hullihen Hall
Newark, DE 19716
Phone: 302-831-2137
Fax: 302-831-2828

DATE:	August 9, 2021
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-25] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	August 9, 2021
EXPIRATION DATE:	July 24, 2022
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)* Subpart D Determination- 45 CFR 46.404

Thank you for your Continuing Review/Progress Report submission to the University of Delaware Institutional Review Board (UD IRB). The UD IRB has reviewed and APPROVED the proposed research and submitted documents via Expedited Review in compliance with the pertinent federal regulations.

As the Principal Investigator for this study, you are responsible for and agree that:

- All research must be conducted in accordance with the protocol and all other study forms as approved in this submission. Any revisions to the approved study procedures or documents must be reviewed and approved by the IRB prior to their implementation. Please use the UD amendment form to request the review of any changes to approved study procedures or documents.
- Informed consent is a process that must allow prospective participants sufficient opportunity to discuss and consider whether to participate. IRB-approved and stamped consent documents must be used when enrolling participants and a written copy shall be given to the person signing the informed consent form.
- Unanticipated problems, serious adverse events involving risk to participants, and all non-compliance issues must be reported to this office in a timely fashion according with the UD requirements for reportable events. All sponsor reporting requirements must also be followed.

Oversight of this study by the UD IRB REQUIRES the submission of a CONTINUING REVIEW seeking the renewal of this IRB approval, which will expire on July 24, 2022. A continuing review/progress report form and up-to-date copies of the protocol form and all other approved study materials must be submitted to the UD IRB at least 45 days prior to the expiration date to allow for the required IRB review of that report.

If you have any questions, please contact the UD IRB Office at (302) 831-2137 or via email at hsrb-research@udel.edu. Please include the study title and reference number in all correspondence with this office.