LANGUAGE DOCUMENTATION & CONSERVATION

Vol. XX (20XX), pp. XXX–XXX
http://nflrc.hawaii.edu/ldc
http://hdl.handle.net/10125/XXXXX
Revised Version Received: Day Month Year

# Describing definites and indefinites

Virginia Dawson[1] and Peter Jenks[2]
[1] *Western Washington University and* [2] *UC Berkeley*

## 1. Introduction

This paper is written to assist native speakers of understudied languages, and linguists collaborating with native speakers of such languages, to understand and describe the sometimes complicated meanings of articles and other markers of definiteness and indefiniteness which occur in human language. The meanings of nominal expressions in languages vary along number of dimensions. By marking distinctions in definiteness and indefiniteness, languages equip their speakers with tools to introduce, identify, and track individuals both in the real world and in the context of a conversation. This semantic function seems to be part of the expressive power of all natural languages, and understanding how languages mark definiteness and indefiniteness is an important part of providing a clear and accurate grammatical description of the language.

While they are central to linguistic meaning, definiteness and indefiniteness have the reputation of being difficult to study, especially using elicitation. These difficulties arise due to a number of factors, central among them the problem of equivalence in translation (Deal, 2015) and the fact that systems of (in)definiteness across languages vary significantly both in inventory and the kinds of distinctions made (Haspelmath 1997; Becker 2021). These complicating factors can be seen, for instance, in cases where there are 'too many translations', such as in the Thai example in (1), where absent a broader context the bare noun in object position can be translated into English as definite or indefinite.[1]

(1) Chăn hěn [**măa**]
1SG see dog
'I saw a/the dog(s).' (Thai)

The converse problem of 'not enough translations' can also arise, as illustrated with the Tiwa indefinites in (2), where multiple phrases can be translated with the indefinite *someone*. These two indefinites are distinguished in what they signal about the context – both in terms of the speaker's epistemic state and in how many individuals could be the one who came – but this difference is masked in the English translation, as English *someone* does not encode any of this information directly.

[1] Throughout this paper, we will indicate the noun phrase under discussion in brackets, while the specific marker of definiteness or indefiniteness will be bold; in the case of bare nouns, this will be the noun itself. This annotation has been added to the data which we cite.



E-ISSN 1934-5275

(2) 'Someone came.' (Tiwa, Dawson 2020:96-97)

a. [**Shar-khí**] phi-ga.
who-INDEF come-PFV

b. [**Shar-pha**] phi-ga.
who-INDEF come-PFV

In order to address such problems of translational equivalency, many researchers have focused on textual data in documenting and analyzing (in)definiteness (e.g. Gundel et al. 1993). Such connected discourses are essential in any documentation of (in)definiteness and provide a ready means of distinguishing old vs. new information (often encoded as definite vs. indefinite, respectively) and other relevant relevant contextual clues.

While we affirm that connected discourse provides essential information in the documentation of (in)definites (and many other domains), we follow Matthewson (2004) in noting that textual data alone may not provide enough information to adequately capture the full range of semantic and pragmatic detail which characterizes a particular kind of nominal. In particular, the texts collected by a researcher may simply lack the relevant contextual information to distinguish two expressions. For instance, the two Tiwa indefinite articles in (2) could be used in the same discourse contexts, despite their semantic differences, with the only difference being the speaker's internal knowledge about how the person may or may not be identified. This information may be otherwise absent from the text itself and from any translation, masking a key semantic distinction. Furthermore, certain kinds of indefinite or definite articles may be less likely to occur in the kinds of text corpuses traditionally compiled by linguists, just as any lexical item might by chance be omitted from a particular sampling of language. Finally, texts can only provide positive evidence for a particular analysis; the absence of a linguistic expression in a particular context in a corpus cannot be taken as conclusive evidence that that such expressions are unnatural in that context. Targeted elicitation provides a means of ascertaining whether this is the case via native speaker judgment.

In this paper we offer a number of tools to complement text-based approaches to documenting definite and indefinite markers. The central tool is an elicitation guide through diagnostic sentences whichs target the attested distinctions in meaning in these domains. In laying out these tools, we hope to highlight the ways which indefinite and definite markers differ in their semantic and pragmatic interpretation across languages, and to provide a systematic way for researchers to explore the semantic possibilities within each of these domains. While we base our elicitation strategies on the emerging typology of (in)definiteness cross-linguistically, work on cross-linguistic semantics (e.g. Von Fintel and Matthewson 2008) has taught us that there is no good reason to assume that any kind of nominal meaning—whether definite or indefinite—is drawn from a known set of universal meanings, that any particular meaning need be expressed in any given language, or that the range of meanings detailed in this article is complete. As a result, the approach adopted in this article is to identify attested semantic categories of definiteness and indefiniteness and to provide diagnostics for those categories. Any nominal expression might then be found to span

multiple meanings, or be restricted to just one. Finer distinctions, or the absence of any distinction, are all open possibilities. The goal is that by bringing the tools in this article to bear on a wider range of languages, researchers can engage in detailed and rigorous semantic documentation and description of these notoriously difficult categories. This in turn, we hope, will continue to expand our understanding of how definiteness and indefiniteness are expressed across languages.

The general structure of this paper is as follows. Section 2 provides a basic introduction to some relevant terms and types of semantic distinctions that are used in this paper. Section 3 then provides an overview of the morphological realization of definiteness, and diagnostic tests and semantic or pragmatic environments for different classes of definite expressions, including a discussion of pronominal distinctions. Section 4 provides an overview of the morphological realization of indefiniteness, and similarly introduces diagnostic tests and semantic or pragmatic environments for different subtypes of indefinite expressions. We offer concluding remarks in Section 5.

**2. (In)definiteness and nominal interpretations** This section defines some terms which will be useful in understanding the following discussion. In particular, we provide a loose definition of definite vs. indefinite, and discuss the contrast between discourse reference, semantic reference and quantification. We also provide a brief overview of other interpretations of noun phrases, which we will not deal with in detail in this paper.

Providing a formal definition of definite vs. indefinite is a complicated task, given the broad disagreement among scholars on the precise nature of these elements, as well as the plain fact that there is a great deal of variation in the core semantics of these expressions. As such, we adopt a definition based on how these elements are used in context. In particular, we adopt the definition in (3).

(3) a. DEFINITE: An NP is definite if it must be used in a context where its intended reference is recoverable.
b. INDEFINITE: An NP is indefinite if it can introduce a new discourse referent whose identity need not be recoverable from the context.

The essence of this definition tracks on the fact that when a speaker uses a definite expression, the addressee can establish what that expression is intended to refer to, a property Lyons (1999) dubs identifiability. In contrast, indefinites can be used in the absence of this identifying contextual information. Consequently, a core use of indefinites is to introduce a new discourse referent whose identity has not been contextually established.

An important distinction needs to be made between different sorts of reference, particularly between discourse reference and semantic reference (Heim, 1982). DISCOURSE REFERENCE has to do with the ability of certain NPs to introduce new entities into a discourse and then the ability of other NPs to refer back to those entities that were introduced earlier. Consider the following mini-discourse.

(4) Did you hear what happened to me yesterday? I had an encounter with [**a** fifteen-foot alligator]$_i$ and [**a** snapping turtle]$_j$. [**The** alligator]$_i$ ran right at me. I managed to dodge [**it**]$_i$, but [**the** snapping turtle]$_j$ nipped my toe!

The two indefinites *a fifteen-foot alligator* and *a snapping turtle* introduce discourse referents—i.e., things that are being talked about—into the discourse. These discourse referents can be subsequently referred back to in English using definite NPs and/or pronouns. SEMANTIC REFERENCE, in contrast, has to do with the core semantic value of the expression itself, rather than how it behaves in a broader discourse. Referential expressions, in this semantic sense, are those for which their core meaning is a specific entity (e.g. a particular person, place, or thing). Speakers use referential expressions to pick out these entities – i.e. to refer. As we will see in the next section, definite NPs are usually referential in this semantic sense. Other referential expressions include proper names and personal pronouns. It is also likely that at least some indefinite expressions are semantically referential, as we will see.

However, not all NPs are referential. Most relevantly for us, many NPs are instead QUANTIFICATIONAL. Rather than pointing directly to a specific entity, quantificational NPs express a relation, often proportional, between the NP property (i.e. the noun and its modifiers) and the rest of the sentence (Barwise and Cooper, 1981). There are many different relations expressed by quantifiers, including the universal quantification expressed in English by *all* and *every*.[2] Most relevant for our purposes is the existential quantification expressed by many indefinites. Consider the sentence in (5), with the indefinite *a cat*.

(5) [ **A** cat ] stole my sock.

This sentence expresses an existential claim: there is some cat that stole the speaker's sock, or, alternatively, such a cat exists. The speaker does not directly refer to a specific cat, but instead existentially quantifies over cats. As a consequence, this sentence will be true so long as some cat—any cat—stole her sock. Note that while the speaker may have a particular cat in mind in using this sentence, she does not refer directly to it, and consequently her utterance will be true even if she is mistaken about which stole her sock.[3] Since indefinites that express existential quantification are not referential, we cannot sensibly talk about the individual they refer to. However, we can still talk about the individual which might make the existential claim true, which, in the case of (5) is whichever cat happened to steal the speaker's sock. This individual is known as the "witness" to the indefinite, as it witnesses the existential claim. The set of potential witnesses to the claim are known as the domain of quantification. While different indefinite NPs may have very different core semantics—some referential, others quantificational—they are all used to introduce discourse referents. Through-

[2]Note that while (in)definiteness is typically marked in an article system, quantifiers can be either D-quantifiers, i.e. part of the determiner system, or A-quantifiers, clausal modifiers of some kind, such as auxiliaries or adverbs (see Partee 1995, and many of the contributions to Bach et al. 1995). We only discuss quantification insofar as its part of the system of indefiniteness in many languages.

[3]This contrasts starkly with referential expressions like *that cat* or *Fluffy*, which would render the sentence untrue if some other cat was the culprit.

out this paper, we will be careful to distinguish discourse referents from semantic referents and witnesses to existential quantification.

Before we dive into the details of definiteness and indefiniteness, there are three other interpretations of NPs beyond referential and quantificational, which we will briefly characterize for completeness. These are predicative NPs, generics, and kinds. First, PREDICATIVE interpretations of NPs are those which occur as the predicate of copular sentences and which attribute a nominal description to the subject. For example, the noun phrases *an excellent teacher* and *the tallest woman in the room* are both predicative in the following sentences:

(6) a. Fatimah is [**an** excellent teacher].
b. Fatimah is [**the** tallest woman in the room].

In English, both definite and indefinite noun phrases can be predicative; *the tallest woman in the room* describes a unique individual who can be identified but also is an attribute that can be ascribed to someone, as in this sentence. In many languages, predicative noun phrases have distinct case patterns, are bare noun phrases or lack articles, or show verbal inflection (Hengeveld, 1992).

Second, GENERIC NOUN PHRASES are those which occur in characterizing sentences, which state a general attribute of a group or type of entity (Carlson and Pelletier, 1995). The English bare plurals *marsupials*, *dogs*, and *people* in the following sentences are all generic because they describe the relevant group in general terms.

(7) a. [ **Marsupials** ] have pouches.
b. [ **Dogs** ] make [ **people** ] happy.

Generics typically allow exceptions; for example, male marsupials do not have pouches, and there are both unpleasant dogs which cause unhappiness and grouchy people who can't stand dogs. Generics are bare nouns in many languages, but they can also be marked definite, particularly if the language is one that requires articles on all NPs. When bare, generics are typically plural or number neutral if a language makes number distinctions. The generic sentences which contain them are typically restricted to a kind of imperfective aspect or irrealis mood rather than referring to a specific event or state.

Finally, KIND-LEVEL noun phrases are those which say something about an entire kind of entity, such as a species or type of technology, rather than generalizing over the individuals within the larger category (Chierchia, 1998). Kind-level reference is typically restricted to particular predicates such as *discover*, *invent*, or *be common*. Some examples are given in (8).

(8) a. [ **Noodles** ] were invented in Ancient China.
b. [ **The** hoopoe ] is common in much of Europe and Africa.

The bare noun *noodles* in (8a) refers to the technology of extruding and boiling flour as a whole, whereas the definite kind term *the hoopoe* in (8b) refers a species of bird. These kind-level claims do not admit exceptions; they are blanket statements about

an entire type. Like generics, most documented cases of kinds are marked with bare nouns or as definite.

In the remainder of this paper we set aside these additional NP interpretations and turn to the details of definite and indefinite NPs as they introduce and track discourse referents. We only discuss quantification insofar as it is relevant for indefinites. Readers who are interested in learning more about quantification more broadly across languages are referred to the many book-length discussions of the topic, including Bach et al. 1995 and Matthewson 2008.

**3. Definiteness** As it describes linguistic expressions rather than a category of meaning, the term 'definite' has been applied to a large class of nominal expressions including pronouns, proper nouns, demonstratives, and NPs with definite articles, often called "definite descriptions" (e.g. Abbott 2004). What these definite expressions share is that their use implies that their reference is somehow recoverable from the context, the definition for definiteness provided in (3). What distinguishes these definite expressions is the kind of contextual information they rely on to help hearers identify the intended referent.

This section describes how definiteness is marked in different languages (section 3.1), then an general overview of definiteness distinctions is provided in section 3.2. The bulk of the section then introduces a series of diagnostics, based on elicitation and translation tasks, which allow various semantic classes of definite expressions to be identified and distinguished. Section 3.2.1 introduces diagnostics for unique definiteness, tools for identifying anaphoric definites are provided in section 3.2.2, and section 3.2.3 discusses properties of a distinct class of salient definites as well, which are still poorly understood. Section 3.2.4 discusses ways of identifying exophoric demonstratives, i.e., demonstratives which are always semantically referential and can typically be accompanied by pointing, as well as some semantic distinctions which exophoric definiteness can utilize. Section 3.3 shows how the same diagnostics for definiteness distinctions can be applied to pronominal distinctions as well. Proper names are not discussed in any detail.

**3.1 Markers of definiteness** Definiteness is typically marked by a determiner of some kind. Such determiners can either be phonologically independent words or bound affixes. An example of a definite marker which is a phonological word can be seen in the Akan (Kwa) definite article *nó*. Definite articles often contrasts with a morphologically free indefinite article, such as Akan *bí*, both of which occur at the right edge of noun phrases.

(9) a. Mɔ̀tɔ́-ɔ̀ [èkùtú **bí**].
1sg.subj-buy-pst orange spec.indef
'I bought (a certain) orange.'

b. [Èkùtù **nɔ̀**] yɛ̀ dɛ̀w dɛ́ɛ̀
orange def.fam be nice so
'The orange was so nice.' (Akan, Arkoh 2011, p. 52)

An example of a affixal definite article is found in Marka-Dafing (Mande), where the definite enclitic *=ú* phonologically fuses with the final vowel of the preceding noun:

**Table 1.** Phonological realization of the Marka-Dafing definite article (partial)

| | N | | N=DEF | |
|---|---|---|---|---|
| a. | jɛ̀ | 'fish' | jɛ́=ɛ́ | 'the fish' |
| b. | lù | 'house' | lú=ú | 'the house' |
| c. | ljɛ́ | 'pig' | ljɛ́=ꜜɛ́ | 'the pig' |

Articles can either be syntactically independent *phrasal articles*, occurring at the periphery of the noun phrase regardless of the presence of a modifier, or *linking articles*, either occurring only in the presence of modifiers or taking distinct morphological shapes in the presence of modifiers (terms form Himmelmann 2001). An example of a phrasal article can be seen in Marka-Dafing below. Although the definite marker is a clitic, as it is phonologically dependent on the preceding word, it always occurs after the last modifier in the noun phrase.

(10) a. músó mɔ̀ sábá
woman CLF 3
'three women'

b. músó mɔ̀ sábɔ =ꜜɔ́
woman CLF 3 =DEF
'the three women'

c. sàmà dʒɛ̀ːní **=ɪ́**
elephant small =DEF
'the small elephant' (Marka-Dafing, notes from elicitation with Rassidatou Konate)

An example of a linking article comes from Amharic, where the definite article occurs as a suffix on nouns when there is no modifier. However, the definite article also suffixes to adjectives, which precede the noun, and optionally repeats when there are multiple adjectives, though its first occurrence is obligatory (Kramer, 2010).

(11) a. bet-**u**
house-DEF
'the house' (Amharic, Kramer 2010, p. 197)

b. tɨllɨk'-**u** t'ɨk'ur(-**u**) bet
house-DEF beautiful(-DEF) house
'the big black house' (Amharic, Kramer 2010, p. 197)

It is not clear if these different means of definite marking have any meaningful correlation with the different types of definites described below, at least in languages with a single means of marking definiteness. For example, Šereikaité (2019) demonstrates that the long-form adjective in Lithuanian—which looks like a linking article—is a strong or anaphoric definite rather than a unique definite marker. On the other

hand, the Amharic definite seems to mark uniqueness. But both the Lithuanian and Amharic definite markers are affixal linking articles. So researchers should not necessarily assume that any particular form of marking definiteness should entail a particular type of definiteness. However, exophoric demonstratives and anaphoric markers are more likely to be morphologically free, while unique definite articles seem more likely to be clitics or affixes.

Different syntactic patterns can mark definiteness as well. For example, in Bangla, a noun (and modifying adjectives) can either occur following a number-classifier unit or before it. When the numeral and adjectives occur after the classifier, the sentence can be interpreted indefinitely. However, when they occur before the numeral, the noun phrase must be interpreted definitely (Dayal 2012):

(12) a. du ʈo **lal boi**
two CL red book
'Two red books'
b. **lal boi** du ʈo
red book two CL
'The two red books' (Dayal 2012, p. 203; translation inferred from text)

Similarly, in Mandarin the position of modifiers relative to a numeral-classifier constituent affect whether the expression is referential (Zhang, 2006). This effect is illustrated in the examples below.

(13) a. lai-le [san ge **dai yanjing de** xuesheng]
come-PRF three CL wear glasses DE student
'Here come three students who wear glasses.'
b. * lai-le [**dai yanjing de** san ge xuesheng]
come-PRF wear glasses DE three CL student

(Mandarin, Zhang, 2006, p. 5-6)

Certain verbs of appearance and disappearance in Mandarin such as 'come' above allow the single argument of that verb to occur before them or after them. In such cases NPs occurring after the verb must be non-referential and indefinite. In such cases, modifiers such as possessors and relative clauses can only occur after the classifier and before the noun. When the modifier occurs before the numeral, the noun phrase as a whole is referential.

Word order asymmetries are often associated with distinctions in (in)definiteness across languages; an open research question is exactly what subtypes of definiteness or indefiniteness can be sensitive to word order effects like those illustrated above in Bangla and Mandarin.

An important typological distinction among languages is that only some languages allow bare nouns, without an article or determiner. When available, bare nouns can be an important part of the pattern of (in)definiteness marking. For example, in Marka-Dafing, definite noun phrases require a definite article or demonstrative, whereas only indefinite noun phrases can be bare nouns. The same is true in Cantonese, where classifiers mark definiteness but bare nouns and numerals are indefinite

(Cheng and Sybesma, 1999). In bare noun languages like Hindi (Dayal, 2004) and Mandarin (Cheng and Sybesma, 1999), bare nouns allow both definite and indefinite interpretations. Finer-grained distinctions have been noted in bare noun languages, as bare nouns generally pattern with unique definites specifically in Thai (Jenks, 2015) and Mandarin (Jenks, 2018).

Argument marking asymmetries can be sensitive to definiteness distinctions. For example, languages with differential object marking (Aissen, 2003) often use the presence or absence of accusative case to indicate definiteness or referentiality, as in Turkish (Enç, 1991a; Von Heusinger and Kornfilt, 2005). Finally, languages sometimes mark contrasts in definiteness with the addition of a marker on the verb, for example, with the addition of a pronominal object marker, as in cases of Bulgarian clitic doubling (Rudin, 1997) or Hungarian objective agreement, which tracks something like definiteness (Coppock and Wechsler, 2012). In such cases, the diagnostics outlined in this paper could help identify the type of definiteness being marked.

**3.2 Subtypes of definiteness** The following categories of definiteness are associated with dedicated definite markers in different languages. They seem to constitute core conceptual categories for definiteness.[4]

(14) a. **Unique definiteness**: Definite reference licensed by contexts where only one individual matches the descriptive content of the noun phrase (e.g., *the bird* when there is just one bird in a room), or the speaker is referring to the entire group of individuals matching the descriptive content of the noun phrase (e.g., *the birds* when there are several birds in a room).

b. **Anaphoric definiteness**: Definite forms which are always discourse referential, typically requiring the explicit previous mention of a particular individual or group in the conversation ('familiar definites' in Schwarz 2009, a subtype of 'endophoric deixis' in Diessel 1999).

c. **Shared knowledge definiteness**: Definite forms which are licensed by shared knowledge between speaker and hearer. The shared knowledge might be licensed by general background past knowledge (**recognitional definiteness**), reference to a particular point in the past where this knowledge was last established (**temporal definiteness**), or reference to a salient referent in the immediate context of utterance (**salient definiteness**) (cf. Barlew, 2014; Becker, 2021).

d. **Exophoric definiteness**: Definite forms which require an associated pointing motion or gesture, or which include a spatial, personal, or visual dimension of reference.

Exophoric, anaphoric, and unique definites have been proposed to form a scale or semantic map (cf. Ahn, 2017; Becker, 2021), where exophoric deixis almost always requires a demonstrative, unique definites almost always requires a definite article if one is available, and anaphoric definites are typically expressed with either demonstratives or definite articles in different languages.

[4]The larger category of 'shared knowledge definites' is being used here for the first time to include the other categories. See section 3.2.3 for discussion.

(15) Exophoric > Anaphoric > Unique

The place of shared knowledge definites is less clear, but they seem to have considerable overlap with anaphoric definites, as we will see below, and sometimes require more marking than exophoric definites.

The scale in (15) has an explanation in the historical development of definite articles, which typically originate from exophoric demonstratives via anaphoric demonstratives (Greenberg, 1978). This scale in is also tied to complexity of form: unique definites are most likely to be bare or have phonologically weak articles, whereas anaphoric and then exophoric definites are increasingly likely to be overtly marked (Jenks, 2018), with anaphoric definites often built on unique definiteness (Royer, to appear; Jenks and Konate, to appear). Exophoric demonstratives often bear stress and permit contrastive uses.

While many languages have determiners which can be used only anaphorically or exophorically, there are languages, such as Korean (Ahn, 2017), Limbum (Becker, 2021), and Marka-Dafing (Jenks and Konate, to appear) which have distinct ways of marking unique, anaphoric, and exophoric definiteness:

(16) a. músó$^{!}$ =**ó**
woman DEF
'the woman'

b. **wó**$^{!}$ músó$^{!}$ =**ó**
DEM:ANAPH woman DEF
'the woman (that we were talking about)'

c. músó $^{!}$**míi**
woman DEM:EXO
'this/that woman' (pointing)

The fact that different definite subtypes can be marked in different ways highlights the importance of attending to the relevant distinctions. The following sections both offer the contexts associated with each of the categories of definiteness outlined above as well as some semantic tests for clearly identifying a particular definite as belonging to one of these categories.

One question that may have no good answer is whether determiners marking each of the categories above should be called demonstratives or definite articles. Becker (2021) proposes that a determiner is a demonstrative if it is possible in exophoric contexts; exclusively anaphoric definite markers would be articles under this definition. Under this definition, languages like Thai or Mandarin that use the same determiner—a 'demonstrative'—in both exophoric and anaphoric contexts cannot be said to have a 'definite article.' The best that can be said, then, is that unique definite determiners have traditionally been called articles, exophoric determiners are traditionally called demonstratives, and that anaphoric determiners have been called both articles and demonstratives.

**3.2.1 Unique definites** Unique definite noun phrases impose a uniqueness condition on the referent in some context. The notion of uniqueness is really only appropriate for singular referents; the plural counterpart of uniqueness is maximality, i.e., reference to the total group of individuals in the context that match the description of the NP. For example, in the following sentences, there should be one and only one cat in the context to use (17a), while the plural noun phrase *the cats* will refer to all of the cats in the context in (17b).

(17) a. Fatimah saw the cat.
b. Fatimah saw the cats.

Uniqueness and maximality have together been grouped under the label 'inclusiveness' by Hawkins (1978, 1991), by which he means that the referent of a definite expression includes all of the available referents in the context .

This section lays out four tests for identifying unique definites: consistency, contextual uniqueness, part-whole bridging, and situation-dependent covariation. These tests collectively identify unique definites; no single test should be seen as conclusive on its own.

First, unique definite show what Löbner (1985) terms *consistency*, that is, they reliably refer to the same individual in the same context. To show a noun phrase shows consistency, the target noun phrase is used as the subject of a predicate and its negation. If a contradiction arises, the definite marker may be a unique definite. If no contradiction arises, the definite marker must be a demonstrative or indefinite marker:

(18) a. #[**The** boy] is sleeping but [**the** boy] is not sleeping.
b. [**That** boy] is sleeping and [**that** boy] is not sleeping. (English, Dayal 2004, p. 417)

The subject *the boy* in (76) reliably picks out the same unique boy in the context, resulting in a contradiction: the same boy cannot be both sleeping and not sleeping. But two demonstratives can pick out different individuals in the same context, so no contradiction results.

Dayal (2004) shows that consistency can be used to show that Hindi demonstratives are not definite markers:

(19) [**vo** laRkaa] so rahaa hai lekin [**vo** laRkaa] nahiiN so rahaa hai
that boy is sleeping but that boy not is sleeping
'That boy is sleeping but that boy is not sleeping' (Hindi, Dayal 2004, p. 417)

While a unique definiteness must show consistency, behaving like English *the*, we will see in the following section that anaphoric definites also show consistency (Moroney, 2019a), making the diagnostic above insufficient on its own to diagnose a unique definite.

The second test for uniqueness involves setting up a context where uniqueness/-maximality is established, and determining whether a definite marker must be used

in such cases. There are two relevant kinds of contexts for unique definites, as established by Hawkins (1978): *immediate situation definites*, in which the uniqueness is due to a specific discourse or real-word context, or 'larger situation definites,' which draw on general shared knowledge of the speaker and hearer that particular types of entities, like the sun or moon, are unique.

We begin with some examples of immediate situation definites: if we are in a room with a single dog, or if we are in a house with a single baby, it is usually obligatory to use a unique definite article to describe the dog or baby for English speakers. To test immediate situation uses, it is best to establish a context where there is no prior mention of the relevant entity in the discourse, which might license an anaphoric definite. Examples are provided below in Turoyo (Semitic: Turkey) (20) and Marka-Dafing (Mande: Burkina Faso) (21).

(20) [**'u** kalbo] carša kokoyu
DEF dog tooth hurt
'The dog has a toothache' (Turoyo, Yifrach & Coppock, ex. 12)

(21) à:ká: [dé **=é**] kùnùn wà
NEG.IMP child =DEF wake.up NEG
'Don't wake up the baby!' (Marka-Dafing, Jenks & Konate, ex. 48)

In these kinds of examples, contexts should involving a specific time, place, and individual, and should contain claims which are not generally true. For example, 'the dog has a toothache' is a useful example for identifying an immediate situation because dogs do not generally have toothaches. If a speaker is claiming that 'dogs like meat' or 'dogs are barking', it might be unclear whether the speaker is making a general claim or referring to a specific situation.

Larger situation definites involve reference to objects which are known to be unique by all ordinary language users, Useful examples include natural objects such as the sun, moon, or ocean, unique positions of leadership such as mayor, chief, president, or principal, and local landmarks such as temples, rivers, or markets. Some general definites, particularly locations like schools and markets, often are used as weak definites, described below, so they should be used with caution or with unusual verbs besides 'go' or 'attend.' The use of a definite article in a larger situation definite is illustrated again for Turoyo in (22) and Marka-Dafing in (23):

(22) Armstrong wa 'u barnošo qamoyo d-faer l-[**u** sahro]
Armstrong COP.PST DEF person first COMP-fly to-DEF moon
'Armstrong was the first person to fly to the moon.' (Turoyo, Yifrach & Coppock, ex. 14)

(23) [té =**!é**], káŋ fàrì
sun =DEF COP.PRES intense
'The sun is intense.'(Marka-Dafing, Jenks & Konate, ex. 46)

In languages with full-fledged unique definiteness marking, articles are usually required in larger situation definites.

One caveat is that certain semantic subclasses of nouns sometimes do not occur with unique definite articles. This is especially true for titles, such as 'president' or 'principle,' which can take on uses more akin to proper names. Different classes of nouns—particularly related to the animacy hierarchy, i.e., kin, humans, and animates—should be tested to determine whether a classifier is necessary. For example, in Bangla, classifiers are typically used to mark unique and anaphoric definites, but with human referents being shown deference or respect, such as lawyers, teachers, or managers, the classifier does not occur, and bare nouns occur instead (Simpson and Biswas, 2016).

If a language lacks definite articles, such as in Moro (Kordofanian: Sudan) (24) or Cuevas Mixtec (Otomanguean: Mexico) (25), bare nouns will typically be used in unique definite contexts. This is illustrated for immediate situation definites below; larger situation definites behave the same:

(24) *Context:* A picture is shown with a single bird in a single tree.
[**ugɜfiə**] g-a-w-ó ík-[**ugi**]
bird CLg-RTC-be.loc-PFV LOC-tree
'The bird is in the tree.' (Moro, elicitation with Angelo Naser)

(25) *Context*: A family's dog has gone missing for a week. A relative enters their house one day to find them cheerful and then proceeds to ask why they are suddenly happy.
ìndyīkókōō [(**#tyíí**) tyìnā].
return.COMPL CL dog
'The dog came home!' (Cuevas Mixtec, Cisneros 2018, p. 2)

The context set up by Cisneros in (25) is an good test for a uniqueness definite because in it, there is a unique dog but there is no previous mention of that dog. The fact that a classifier in Cuevas Mixtec was not available in this context provides evidence that classifiers do not mark unique definiteness; Cisneros shows in the same paper that the classifier marks an anaphoric definiteness.

The third test to identify unique definiteness is *part-whole bridging* (the term is from Schwarz 2009), where an entity is introduced and then some unique part of that entity is referred to. Examples are given below for Fering (Germanic: Germany) and Turoyo:

(26) a. Wi foon a sark uun a maden faan't taarep.
We found a church in the middle of.the villiage.
'We found the church in the middle of the village.'

b. [**A** törem] stän wat skiaf.
'the$_{\text{weak}}$ tower stood a.little crooked
'The tower was a little crooked.' (Fering, Ebert 1971, p. 118, as cited in Schwarz 2009)

(27) a. hze-lan cito b-falge d-i krito.
saw-to.us church in-middle of-DEF village
'We found a church in the middle of the village.'

b. [**'u** burgo] ëšmto cwiyo wa
DEF tower a little bit crooked COP.PST
'The tower was a little crooked.' (Turoyo, Yifrach & Coppock, ex. 15) )

The definite articles in (26b) and (27b) are licensed by what is presumably a culturally salient assumption that churches have just one tower. Any unique part could be similarly licensed by a previously mentioned whole as long as common knowledge is sufficient to establish uniqueness, for example, the trunk of a tree, the tail of an animal, or the roof of a house.

The last test for a unique definite is that it allows *situation-dependent covarying readings* (Schwarz, 2009). In such cases, a single unique definite is able to pick out different individuals in different contextually specified situations, as illustrated below in English:

(28) a. In every house, we looked up [**the** chimney].
b. In every village, we visited [**the** chief].
c. In every classroom, [**the** teacher] was talking.

Covarying readings of unique definites are available in these sentences without any prior mention of chimneys, chiefs, or teachers. These types of contexts are similar to part-whole bridging in that the topic refers to a place or entity which is known to have some correspondingly unique entity, e.g. the roof of a house, the chief of a village, or the teacher of a classroom, hence licencing reference to that entity via a unique definite.

The relevant reading of the sentence must be one where the targeted definite NP picks out a different individual in each of the contexts, as is the case, for example, in the German weak definite in (29), as well as with the Dafing definite article in (30):

(29) In jeder Stadt, in der unser Zug hielt, wurde mir der Brief vo-[**m** /
In every city in which our train stopped was me the brief from-the$_{\text{UNIQUE}}$
#von [**dem** Bürgermeister] überreicht.
from the$_{\text{ANAPHORIC}}$ mayor handed
'In every city we stopped in, a letter was handed to me from the mayor.' (German, Schwarz, 2009, p. 231)

(30) *Context*: It is well known that chiefs are mean and grumpy people.
zúù vjè sɔ̀ [zúʃí =$^{!}$**í**] kò-màŋ-ʒí mɔ́ =$^{!}$ù yè wà
village-PL every in chief =DEF not.like people =DEF.PL with NEG
'In all the villages, the people don't like the chief.' (Marka Dafing, Jenks and Konate, to appear, ex. 53)

Anaphoric definites do not allow situation-dependent covariation, as shown in the unavailability of the contracted article in German in (29). We will see in the following section that using an anaphoric definite in such a context results in a reading which refers back to a specific individual regardless of the choice of situation. Summarizing the section to this point, then, a definite expression can be diagnosed as definite if it 1) shows consistency effects, 2) is required in a range of contextually unique situations,

and 3) allows situation-dependent covariation. A variant of situation-dependent covariation is the 'president sentences' discussed in the following section (see examples (63-64)).

In languages where bare nouns are prohibited or more restricted in their distribution, unique definite articles also occur on noun phrases which are not referential at all. For example kind-level and generic nouns (discussed in Section 2) typically occur with definite articles in Romance languages such as Italian.

(31) [**Il** cane] e rare.
The dogs are rare
'Dogs are rare.'

(32) [**Il** cane] amano giocare.
The dogs love play.INF
'Dogs love to play.' (Italian, Chierchia, 1998, p. 342)

Dedicated anaphoric, recognitional, or deictic definite markers would not be expected in such contexts, as these markers are always referential.

Relatedly, unique definite articles sometimes occur with nouns which are the objects of common or prototypical activities; such non-referential uses of definite articles are typically labeled *weak definites* in the formal semantics literature (e.g. Carlson et al., 2006):

(33) Weak definites in Standardized English
a. My grandmother would often listen to [**the** radio].
b. Someone needs to open [**the** window].

Weak definites seem to pose a problem for the requirement that there be some contextually unique entity which licences the use of the definite article. In particular, the speaker's grandmother could have listened to any number of radios in (33a), (33b) is fine in a room with two windows, and no contextually unique store needs to be known to the speaker and hearer for (??) to feel natural. Birner and Ward (1994), looking at a very similar set of cases, observe that in such cases the referent is "both undifferentiated and not relevantly differentiable in context" (Kadmon, 1990, see also).

Weak definites can be distinguished from other unique definites in that the definite noun phrase is only really possible with a few specific verbs that describe the common activity (Carlson et al., 2006). For example, the following variants of the sentences above, with the verbs changed, all do seem to refer to some contextually unique entity, and do not allow weak definite readings.

(34) a. My grandmother would often fix the radio.
b. Someone needs to clean the window.
c. Yesterday we called the store.

What is important about weak definites from the perspective of documentation and description is that they might seem to pose potential problems for descriptive generalizations based on uniqueness but in fact are typical environments for unique definite

NPs in many languages. The important takeaway is that when testing for unique definiteness, sentences which describe culturally prototypical or common types of activities should be treated with caution, as these might trigger uses of weak definite uses of definite articles, in which the definite noun phrase does not seem to be referring to any particular entity. It is also possible that a language might not mark definiteness at all in weak definite contexts, particularly if bare nouns are available for generic or kind-level reference.

**3.2.2 Anaphoric definites** Anaphoric definites, also called familiar definites and strong definites by Schwarz (2009), are definite noun phrases which refer back to an individual that is explicitly mentioned earlier in a discourse. While the idea that definiteness might be characterized by discourse oldness or familiarity has been around for some time (Christophersen, 1939; Heim, 1982, 1983), such accounts typically treated definiteness as equivalent to familiarity. Only since the work on German and close relatives by Schwarz (2009, 2013) has it been clear that anaphoric definites are a typologically distinct subtype of definite expression, a finding since replicated in many languages. This section outlines some of the contexts that trigger anaphoric definite uses; see Schwarz (2013) for a similar survey and additional examples; Becker (2021) and Aguilar-Guevara et al. (2019) discuss a number of relevant cases from different languages as well.

Anaphoric definites take different forms in different languages. There are three different patterns. First, some languages have a dedicated definite article or determiner for anaphoric definiteness which contrasts with a unique definite article, including German and Fering (Schwarz, 2009), Lakhota (Rood 1942,O'Gorman 2011), Hausa (Buba, 1997; Jaggar, 2001)—all discussed in Schwarz 2013. Other languages use bare nouns in unique definite contexts which is contrasted with a dedicated anaphoric definite marker, which can take many forms. Languages where it is simply an anaphoric determiner or demonstrative include Runyangkore (Asiimwe, 2014) Malagasy (Paul, 2009), Korean (Ahn, 2017) and American Sign Language (Irani, 2019). The dedicated anaphoric definite marker can also be a pronominal determiner (Tiwa, Dawson 2020), a classifier (Cuevas Mixtec, Cisneros 2019), or a regular demonstrative, as in the case in Thai (Jenks, 2015) and Mandarin (Jenks, 2018). Finally, some languages use a double marking pattern, requiring a unique definite marker along with an additional determiner for anaphoricity. In Chuj (Royer, to appear), the unique-marking classifier is doubled by a general purpose demonstrative which is used in anahoric contexts; in Marka-Dafing (Jenks and Konate, to appear), a regular definite article is doubled by a dedicated anaphoric demonstrative. One absolute universal about unique and anaphoric definite marking is that anaphoric definites never exhibit less morphosyntactic complexity than their unique definite counterparts (Jenks, 2018).

This section presents several tests for identifying anaphoric definite markers, many of them specifically distinguish unique and anaphoric definites. Such diagnostic tests are important because many anaphoric contexts allow both unique and anaphoric definites, so texts alone can be confusing in this regard; we discuss the contexts where

either kind of definite expression might occur. Not all languages pattern the same in all of these contexts; where important differences have been noted in the literature, we try to make note of them.

First, anaphoric definites should pattern with unique definites in showing consistency effects. Moroney (2019b) shows that anaphoric definite uses of demonstratives show consistency effects in English (**??**) and Thai (36), and the example in (37) shows the same for the Korean anaphoric demonstrative *ku*.

(35) There is a child in the next room. #[**That** child] is sleeping but [**that** child] is not sleeping. (English, Moroney, 2019a, p. 7)

(36) Mii dèk khon nì̀ŋ yùu nay hɔ̂ɔŋ thàt pay. #[dèk khon **nán**] nɔ̀ɔn yùu tɛ̀ɛ [dèk
have child CLF one LOC in room next PRT child CLF that sleep IPFV but child
khon **nán**] mâj.dâj nɔɔn yùu.
CLF that NEG sleep IPFV

'There is a child in the next room. #That child is sleeping but that child is not sleeping. (Thai, Moroney, 2019a, p. 7)

(37) Context: If the hearer went to the other side of the room to grab something I asked for, and is turned toward A, and I want B which is behind him:

[**ku**-kess] mal-ko! *[**ku**-kess]! / [**ce**-kess] [pointing at book 2]
ku-thing not-CONJ ku-thing ce-thing

'Not that book, that book!' Ahn (Korean, 2017, ex. 33)

In both English and Thai, the same demonstrative in an exophoric context does not show consistency (see below). Ahn (2017, ex. 33) shows that the anaphoric demonstrative *ku* in Korean also exhibits what seems to be a consistency effect.

A second way to test for anaphoric definites is to see if they are infelicitous—that is, pragmatically or semantically strange, marked '#'—in the context of larger situation definites like 'sun,' 'moon,' or 'president.' While we have seen that such nouns license unique definite articles without prior mention, the use of an anaphoric definite with them is strange with nouns like 'sun' and 'moon', and might trigger the judgment that there must be some particular 'president' that has been mentioned. The first effect is illustrated below for Thai and Cuevas Mixtec, where demonstratives, otherwise used as anaphoric definites in Thai, are impossible, and where Cuevas Mixtec classifiers are similarly impossible.

(38) [duaŋ-can (#duaŋ **nán**)] sàwàaŋ mâak
moon (CLF that) bright very

'The moon is very bright.' (Thai, Jenks 2015, ex. 8

(39) tyà juáàn ndé'é rā [(#**ña**) yóò]
the.SG.M Juan look.IPFV 3.SG.M the.INA moon

'Juan is looking at the face of the moon.' (Cuevas Mixtec, Cisneros 2019, p. 65)

The impossibility of a definite expression in a larger situation definite context constitutes clear evidence for its status as an anaphoric rather than a unique definite.

However, there are languages such as Tiwa (Dawson, 2020), where the pronominal article, which otherwise functions as an anaphoric definite, does occur in larger situation definites:

(40) Táw [(**pe**) chonái] khúp plas-do
today (3SG) moon INTS bright-IPFV
'The moon is very bright tonight.' Comment with *pe*: 'Rare.' (Tiwa, Dawson 2020, p. 56)

The speaker comment shows that bare nouns in Tiwa are still preferred to pronoun-noun sequences, so we can still say that anaphoric definites are dispreferred to unique definites in such contexts.

Another useful way for diagnosing anaphoric definites is to examine their distribution in texts. One clear expectation is that textual occurrences of anaphoric definites should always have a discourse antecedent, as long as they are truly anaphoric definites rather than unique or shared knowledge definites (see section 3.2.3). So one way of confirming the claim that a particular definite expression is anaphoric is simply to identify antecedents for all of its textual occurrences.

However, anaphoric definites are not typically used in every possible anaphoric mention of a noun phrase in texts, in part because pronouns and unique definites often occur anaphorically as well. However, there are several contexts that seem to favor anaphoric definites and demonstratives, including re-activating an previously established topic (Gundel et al., 1993), ambiguity resolution in the case of multiple referents (Himmelmann, 1996), and in "immediate anaphora after first mention" Lichtenberk (1996). The last category refers to the use of anaphoric definites to refer to a new discourse participant with an anaphoric definite in its first anaphoric use, establishing that discourse referent as the current topic.

It is this latter 'new topic' use of anaphoric definites which makes them easy to elicit in our third elicitation-based test, narrative sequences. In such sequences an entity is introduced in one sentence and then immediately referred to in the following sentence; this is illustrated for Marka-Dafing in (41) and for Cuevas Mixtec in (42). In both cases, the first sentence introduces an entity with an indefinite marker, *dò* 'some' in Marka-Dafing (41a), or a bare noun in Cuevas Mixtec (42). The second sentence then refers back to that entity with an anphoric definite marker: *wó* in Marka-Dafing, and the animate classifier *tyí* in Cuevas Mixtec. The narrative sequence is assumed to be the start of a story; it may be useful for speakers to imagine such sentences at the beginning of a story or recounting of events.

(41) a. kúnúŋ músó dò dó-ná m̀m̀ lú kɔ̀ŋ
yesterday woman some enter-PST my house IN
'Yesterday, some woman walked into my house.'

b. ň tí má [ʔ(**wó**) músó =**ó**] ye a-ye wa
1SG PFV NEG ANAPH woman =DEF see before NEG
'I'd never seen that woman before.' (Jenks and Konate, to appear, ex. 52)

(42) isayā'nḯ tyà juáàn kólō
kill.COMPL the.SG.M Juan male.turkey
'Juan killed a turkey.'

a. # káchí nā ñà kú'vì vā kólō
say.IPFV 3.HUM COMP sick.IPFV FOC male.turkey
'They say that a turkey was sick.'

b. káchí nā ñà kú'vì vā [**tyí** kólō]
say.IPFV 3.HUM COMP sick.IPFV FOC the.ANIM male.turkey
'They say that the turkey was sick.' (, p. 68)

The pattern of acceptability is slightly different in Marka-Dafing and Cuevas Mixtec, as the unique definite article is somewhat dispreferred in Marka-Dafing but still possible while it is marked infelicitous in Cuevas Mixtec.

This kind of variation is characteristic of anphoric definites but remains poorly understood. It is certainly true that languages differ in exactly how unique versus anaphoric definites are distributed. One general trend seems to be that in languages with an overt unique article vs. anaphoric article, like German (Schwarz, 2009) or Marka-Dafing (Jenks and Konate, to appear), both uses are generally possible in anaphoric contexts. However, in many languages where an anaphoric definite is in opposition to a unique definite bare noun, the bare noun is rejected in at least some anaphoric contexts, as is the case in Cuevas Mixtec above as well as in languages like Thai (Jenks, 2015), Mandarin (Jenks, 2018), and ASL (Irani, 2019).

Nevertheless, it is still possible to find anaphoric bare nouns in languages with no definite article, as Dayal and Jiang (2020) show for Mandarin. They give the following narrative sequence as one of several examples, where (43b) and (43c) are two potential continuations of (43a):

(43) a. Jiaoshi li zuo zhe [yi ge nansheng] [yi ge nüsheng],
classroom inside sit PROG one Cl boy one Cl girl
'There is a boy and a girl sitting in the classroom...

b. [**nüsheng**] zuo zai [**nansheng**] pangbian.
girl sit DUR boy side
'The girl was sitting next to the boy.'

c. Wo zuotian yudao [#(**na** ge) nansheng].
I yesterday meet that Cl boy
'I met the boy yesterday.' (Dayal and Jiang, 2020, ex. 20)

(43b) allows anaphoric bare nouns (anaphoric demonstratives are also possible) in this context, while (43c), from Jenks (2018), only allows an anaphoric demonstrative. According to Dayal and Jiang, the contextual factor which leads to the availability of anaphoric bare nouns is that the situation where the anaphoric bare noun is used in must be indistinguishable from the situation where its indefinite antecedent is used, at least with respect to the uniqueness of the intended referent. The anaphoric demonstrative in (43c) is required, they suggest, because the situation of 'yesterday' raises the possibility that some other boys were included in the situation, making the unique

definite a less ideal choice. The conclusion from this discussion is that while narrative sequences and texts can be used to identify anaphoric occurrences of anaphoric definites, there is no expectation that only anaphoric definites should occur in such contexts.

Because anaphoric contexts often allow both anaphoric and unique definites, a more reliable way of conclusively distinguishing unique and anaphoric definites is to put the two definite types in contexts which allow situation-dependent covariation, discussed in the previous section (29,30). While unique definites allow covarying interpretations in such contexts, anaphoric definites do not, but instead must be discourse referential, picking out a single individual across situations. This is illustrated for Tiwa (Tibeto-Burman), where the context is built in such a way that the anaphoric definite is infelicitous.

(44) Etha, PM mewa hóng-do
now PM man COP-IPFV
'Now the PM is a man.'

a. Tin-shá-ne tin-a, PM margi hóng-do
day-one-GEN day-DAT, [**PM**] woman COP-IPFV
'One day, the PM will be a woman.'

b. # Tin-shá-ne tin-a, [**pe** PM] margi hóng-do
day-one-GEN day-DAT, 3SG PM woman COP-IPFV
'One day, that PM will be a woman.' (Tiwa, Dawson, 2020, p. 56)

The only reading of (44b), with the pronominal definite marker *pe*, is that the man who happens to be the PM now will be a woman one day. The sentence with a bare noun (44b) allows the situation-dependent covarying reading, which is that one day there will be a female PM. This sentence shows that *pe* is an anaphoric definite marker because it requires discourse reference to the previously mentioned individual in (44). Such sentences are easy to construct and are quite conclusive.

Another test for distinguishing anaphoric definites from unique definites is so-called 'donkey sentences,' where a pronoun or definite expression covaries with a preceding indefinite embedded in a relative clause or conditional (only relative clauses are shown). Such sentences have received intense interest in the formal semantics literature, where the challenge is understanding how the 'donkey anaphor' (bolded below) is able to pick out different donkeys for each owner. In English, such sentences seem to allow pronouns, definite descriptions, and demonstratives (Geach, 1962; Abbott, 2002; Elbourne, 2013):

(45) a. Every man who owns a donkey beats [**it**].
b. Every man who owns a donkey feeds [**the** donkey].
c. Every man who owns a donkey feeds [**that** donkey].

However, not all languages behave like English in allowing unique definites as donkey anaphora; in many languages it has been shown that anaphoric definites but not unique definites are possible as donkey anaphors, shown below in Mandarin and Chuj (Mayan):

(46) a. mei ge [you yi zhi shuiniu de] nongfu dou hui da [**shuiniu**].
every CLF have one CLF buffalo REL farmer all will hit buffalo
'Every farmer that has a buffalo hits buffalo (generally).' (no covarying reading)

b. mei ge [you yi zhi shuiniu de] nongfu dou hui da [**na** zhi shuiniu].
every CLF have one CLF buffalo REL farmer all will hit that CLF BUFFALO
'Every farmer that has [a buffalo]$_i$ hits [that buffalo]$_i$.' (Mandarin, Jenks, 2018, p. 503)

(47) Masanil anima' ix-il-an junjun much, ix-s-mak'-cham [**nok'** much #(**chi'**)]
every person PFV-see-AF INDF.DIST bird, PFV-A3-hit-die CLF bird DEM
heb'.
PL
'Every person that saw a bird, hunted that bird.' (Chuj, Royer, to appear, ex. 20)

In Mandarin, the bare noun, otherwise possible as a unique definite, can only recieve a generic interpretation in donkey sentences (46a); the demonstrative *na* must be used as a donkey anaphor (**??**). Similarly, in the Chuj example in (47), the classifier which marks unique definites is not sufficient to license a donkey anaphoric reading, it must be accompanied by the demonstrative *chi'* which also occurs in anaphoric contexts. What explains the contrast between English-like languages which allow unique definites to occur in such contexts and languages like Mandarin and Chuj remains unclear (see Moroney 2021 for some recent discussion).

One final test which can be used to diagnose anaphoric definiteness, first described in Schwarz (2009), is *producer-product* bridging. In these cases, prior mention of a 'product' licenses an anaphoric definite use of its 'producer' — even without prior mention of that individual. Examples include 'painting-painter,' 'symphony-composer,' 'movie-director' and 'poem-poet.' Since first documented for German and Fering by Schwarz, producer-product bridging licensing anaphoric definites has been documented in a number of other unrelated languages, including Thai and Cuevas Mixtec.

(48) Nít̀ khít wâa klɔɔn bòt nán prɔ́ʔ mâak, mɛ̂ɛ-wâa kháw cà mâj chôɔp
Nit thinks COMP poem CLF that melodious very, although 3P IRR NEG like
náktɛ̀ɛŋklɔɔn #(khon nán).
poet CLF that
'Nit thinks that poem is beautiful, though he doesn't really like the poet.' (Thai, Jenks, 2015)

(49) a. [Tyà Juáàn] ìsyīīn rā [īīn tūtū]
the.SG.M Juan buy.COMPL 3SG.M one book
'Juan bought a book.'

b. [#(tyà) āūtóòr] kúù rā [tyà nūū nūū yūkù ]
the.SG.M author be.IPFV 3SG.M the.SG.M village face mountain
'The author was from San Miguel Cuevas.' (Cuevas Mixtec, Cisneros, 2019, p. 71)

The flip side of this test is that anaphoric definties should be infelicitous in the part-whole bridging contexts described in section 3.2.1.

Before concluding, three additional points should be kept in mind regarding the identification of anaphoric definites. First, the presence or absence of different kinds of modifiers might affect whether a definite be marked anaphorically or not. In Lithuanian (Šereikaité, 2019) and Icelandic (Ingason, 2016), strong definiteness is only clearly marked in the presence of modifiers. Additionally, in German, the anaphoric definite article must be used with restrictive relative clauses (Schwarz, 2009; Wiltschko, 2012).

Second, choice of noun seems to be relevant to how definiteness is marked anaphorically or otherwise. A clear discussion of language-internal variation is provided in Cisneros (2019), who shows that the nouns meaning 'man,' 'woman,' and 'people' require classifiers as definite markers when unique, even though these same definite classifiers only occur anaphorically for other nouns. At the same time, there are a class of nouns such as *tyàxìnì* 'mayor' which have a built-in classifier (*tyà*), and correspondingly allow anaphoric uses even as bare nouns.

Finally, it seems clear that more subtle variation exists in exactly what additional pragmatic conditions are associated with anaphoric definites in different languages. For example, in a language like Marka-Dafing, speakers have a clear intuition that the anaphoric definite marker *wó* is associated with prior mention; this is not the case for Mandarin or Thai, where the demonstratives can be used either anaphorically or exophorically. There is also claimed to be at least one language with an internal distinction in anaphoric definites: Lichtenberk (1996) identifies separate 'recent mention' and 'distant mention' anaphoric definites in To'aba'ita (Oceanic: Solomon Islands). Jenks (2018) makes a similar observation for Cantonese, where distal demonstratives can be added to classifiers, which mark both unique and anaphoric definites, to mark mention at an earlier point in the discourse.

**3.2.3 Shared knowledge: Recognitional, temporal, and salient definites** In addition to unique and anphoric definites, there are a group of definite expressions which seem to be licensed by shared knowledge between between conversational participants. Three subcases have been identified, here we suggest they might be seen as part of a larger category of 'shared knowledge definites.' In the first case, recognitional definites, the definite seems to be referring back to some entity that both the speaker and hearer have shared knowledge of. The second case, temporal demonstratives, further mark remoteness distinctions as to when the referent was last part of the attentional space of the conversational participants. In the third case of salient definites, this entity is at the current center of the speaker and addressee's shared attentional space. We will describe both these definite categories briefly here and highlight that they remain an important topic for future work. In particular, there is virtually no work examining how these categories behave with respect to the diagnostic tests above, nor is there work examining the extent to which 'recognitional,' 'temporal,' and 'salient' definites overlap in their distribution, or whether should be kept distinct.

The term recognitional definite dates back at least to Himmelmann (1996), who used it to describe demonstratives such as the following English example from (Becker, 2021):

(50) Do you remember [**that** cat] (we used to have)?

While English uses a distal demonstrative to mark recognitional definiteness, Becker (2021) identifies several languages which seem to have dedicated recognitional definite articles. Many but not all of the cases Becker describes have a 'reminding' function, although some examples resemble regular anaphoric definites in being used in "immediate anaphora after first mention." Nevertheless, at least some cases described as recognitional definites by Becker do not require prior mention, distinguishing them clearly from anaphoric definites; but serve to remind speakers of a previously discussed or encountered entity.

Temporal demonstratives, attested in Nilo-Saharan languages Ik (Schrock, 2017) and Kipsigis (Kouneli, 2019), are semantically referential expressions which mark not only that the speaker and hearer share knowledge of the referent is familiar to both speaker and hearer but indicate how distantly in the past the referent was last mentioned or encountered. Consider the singular paradigm of temporal demonstratives in Kipsigis, which must occur with the proximal demonstrative, both nominal suffixes (Kouneli, 2019, p. 144). Underlining here indicates vowels are –ATR]kouneli19; the temporal demonstrative suffix is bolded.

(51) a. lɑ̱ɑ̱kwɑ̱ɑ̱-nì-**kɑ̂ɑ̱n**
girl-PROX-PAST1
'this girl from earlier today'

b. lɑ̱ɑ̱kwɑ̱ɑ̱-nì-**kóɔ̱nyɛ̀**
girl-PROX-PAST2.SG
'this girl from yesterday'

c. lɑ̱ɑ̱kwɑ̱ɑ̱-nì-**kíɪ̱nyɛ̀**
girl-PROX-PAST3.SG
'this girl from long ago'

The fact that temporal demonstratives in Kipsigis are built on proximal demonstratives is reminiscent of anaphoric definites and demonstratives in languages like Marka-Dafing or Greek which require definite articles. Temporal demonstratives are not simply special-purpose anaphoric definites; both Kipsigis and Ik have a separate anaphoric demonstratives. The similarity to recognitional definiteness is clear, however, and in the context of Kipsigis data like (51), recognitional definites perhaps could be seen as 'simple past' temporal definites.

Closely related to recognitional and temporal definiteness is salient definiteness, a category only identified in one language, Bulu (Bantu: Cameroon) (Barlew, 2014). In Bulu, salient definiteness is marked by the suffix *-tè*, which overlaps with an anaphoric definite in its distribution but occurs in a wider range of contexts characterized by shared immediate attention, which Barlew dubs salience. Like recognitional definites, Bulu *-t'e* is compatible with anaphoric uses, and like recognitional definites it does

not require prior mention or require pointing. Barlew illustrates the role of shared attention in pairs such as the following, where the first example uses a bare noun as a unique definite to refer to the sun, whereas in the second context, the sun has been made salient to both speaker and addressee by the actions of the speaker.

(52) *Context:* Abondo is sitting on a bus when a man he does not know sits down beside him. The man says:
# [vjǎn **tè**] wɔ́ fài dɔ́n
sun TE $PN_{11}$ shine today
'The sun is bright today.' (intended)

(53) *Context:* Identical to (5a), except that the stranger first opens the window shade on the bus, letting in sunlight.
[vjǎn **tè**] wɔ́ fài dɔ́n
sun TE $PN_{11}$ shine today
'The sun is bright today.' (Bulu, Barlew, 2014, p. 627)

Both speaker and hearer have their attention focused on the bright sunlight in (53), so the salient definite marker *-te* is licensed, even though its referent 'sun' is a larger situation definite; we saw above that anaphoric definites typically do not occur in larger situation definite contexts.

Barlew provides several similar examples; in each, the details of the context make it clear that conversational participants have focused their attention on a single referent, and *-te* to pick out that referent.

Another possible candidate for a salient definite is what Aikhenvald (2008) dubs a "current relevance" demonstrative in Manambu (Ndu: Papua New Guinea). Aikhenwald cites several naturally occurring uses of this demonstrative with larger situation definites (p. 213), and observes that these demonstratives are licensed when the referent is "within the discourse…or within the frame of attention of the participants in a conversation."

If recognitional and temporal definites encode reference to an entity with which speaker and hearer have shared past knowledge, salient definites might be conceived as present or immediate counterpart to the recognitional and temporal categories; indeed, Schrock (2017, pp. 520–521) explicitly identifies a 'non-past' temporal demonstrative identical to the proximal demonstrative. Again, however, it is not clear at all whether these three categories actually share semantic properties in the languages above. For example, how they behave in situation-dependent covarying contexts, whether they show consistency effects.

One possibility worth considering is that at least some of the documented cases of anaphoric definiteness in the literature are some kind of recognitional, temporal, or salient definite, or that these three categories might show significant overlap, particularly since anaphoric uses are often a subset of recognitional ones. For example, the description of Korean *ku* in Ahn (2017) resembles Bulu *-te* in that it is generally licensed by either by joint attention or anaphoric reference, but does not seem to explicitly require prior mention. Similarly, the Tiwa pronominal demosntrative discussed

above could be licensed by salience, explaining its otherwise exceptional ability to occur with larger situation definites, typically unavailable for strictly anaphoric definties. Future work on anaphoric definites should be more careful to identify which anaphoric definites can also be licensed by shared attention. At this point it is unclear whether any languages have separate determiners or demonstratives for salient, recognitional, and anaphoric reference, as there seems to be a fair amount of overlap in their distribution, and recognitional and salient determiners seem to encompass anaphoric uses.

**3.2.4 Demonstratives and exophoric definiteness** The term 'demonstrative' is usually associated with exophoric uses, that is to say, semantic reference to entities via reference to their location in the physical environment of utterance, typically accompanied by a pointing gesture or demonstration. However, the term 'demonstrative' is also often used to describe recognitional and anaphoric determiners, especially when they clearly form part of a larger paradigm of demonstratives including exophoric uses. The typological literature has classified different uses of demonstratives and identifying the factors that influence one or another demonstratives to be used in different contexts in a given language, as well as identifying cases where demonstratives can span multiple uses. Seminal works include Himmelmann (1996), who lays out a widely-adopted typology of demonstrative uses, and Diessel (1999), who discusses the morphology, syntax, and semantics of demonstratives across languages; recent overview articles include Levinson (2018) and Diessel and Coventry (2020). Recent book-length descriptions of demonstratives in a number of languages can be found in Levinson et al. (2018) and Næss et al. (2020), the latter focusing on anaphoric uses of demonstratives.

One important factor in describing demonstratives is determining which demonstratives only allow exophoric uses and which might allow both anaphoric and exophoric uses, the general norm. However, there are languages with exophoric demonstratives that are distinct from anaphoric definites, the latter of which may or may not be a demonstrative. In Marka-Dafing, for example, there is a dedicated exophoric demonstrative *míi* which requires the speaker to point or otherwise gesture to the intended referent when using it.

(54) wúrú ǃmíi
dog D:EXO
'this/that dog' (pointing) (Jenks and Konate, to appear)

The exophoric demonstrative *míi* in Marka-Dafing is not sensitive to distance from the speaker or addressee, it only requires a pointing gesture, and crucially, it cannot be used anaphorically, as Marka-Dafing has both plain definite articles and the dedicated anaphoric definite marker *wó* for anaphoric contexts. Similarly, Korean has a dedicated exophoric demonstrative *ce*, which is distinct from an anaphoric form *ku* (Ahn, 2017).

There are two simple tests to establish whether a particular demonstrative allows exophoric reference. The first test is consistency. While unique or anaphoric uses

of definite markers show the property of consistency, always picking out the same referent in some context, demonstratives that allow exophoric meaning should allow allow two different individuals to be picked out by multiple uses of the same demonstrative in a single sentence as long as those uses are accompanied by pointing or other demonstrations. We repeat the examples below from earlier.

(55) a. # [**The** boy] is sleeping and the boy is not sleeping.
b. [**That** boy] is sleeping and that boy is not. (English)
c. [**vo** laRkaa] so rahaa hai lekin vo laRkaa nahiiN so rahaa hai
that boy is sleeping but that boy not is sleeping
'That boy is sleeping but that boy is not sleeping.' (Hindi, Dayal 2004, p. 417)

Applying the consistency test actually includes the second, more basic test for exophoric reference, which is determining whether a particular demonstrative expression can or must be accompanied by a pointing gesture. As we have already seen, demonstratives across languages show significant differences in whether pointing demonstrations are required or not. Because of this, it is important to include in any description of a demonstrative system which of the available demonstratives can be accompanied by pointing gestures. The flip side of this is that it is also could be important to test demonstratives in anaphoric contexts; if they are not possible in narrative sequences, for example, they may not allow anaphoric uses at all.

One potential pitfall of a pointing test on its own is that some languages have invisibility demonstratives which are only possible if the referent is not in the current visual field of the speaker; speakers might judge pointing to visible referents incompatible with these demonstratives, despite the fact that these demonstratives are exophoric. For example, there is an invisibility demonstrative in Ticuna (isolate; Peru, Brazil, Ecuador) which must be used even for objects that are close to the speaker as long as they are hidden from view (Skilton, 2019). The same demonstrative is also used for visible objects within reach of the hearer but out of reach of the speaker. Last, the invisibility demonstrative is also used anaphorically. Skilton does not check consistency for these demonstratives but we would expect that the demonstrative in question would not show consistency effects at least in its 'near hearer' uses; it is unknown whether the invisibility uses show consistency or not. Levinson (2018) surveys several other cases of languages with visibility distinctions.

Beyond the two simple tests above which serve to diagnose a demonstrative as allowing exophoric uses, a detailed questionnaire proposed by Wilkins (1999) has been widely adopted for documenting demonstrative distinctions.[5] This paper will not survey the many spatial distinctions that can be made by demonstratives, as there is extensive discussion of this issue in the literature cited above.

**3.3 Pronominal distinctions and definiteness** Contrasts between different third person definite pronouns, e.g. the contrast between English *it*, a personal pronoun, and

[5]This questionnaire is available online at `https://pure.mpg.de/rest/items/item_2573775/component/file_2573783/content`

*that*, a demonstrative pronoun, show parallels to the definiteness distinctions discussed in the earlier sections. For example, demonstrative pronouns allow exophoric reference, while unstressed pronouns such as *it* do not, but are licensed by something closer to salience as defined above (Roberts, 2004). Many languages also have a third 'series' of pronouns: null pronouns, which in some languages must be anaphoric or unique.[6]

There have been relatively few wide-scale typologies examining pronominal distinctions, with a notable early exception being Déchaine and Wiltschko (2002). More recent literature—including Patel-Grosz and Grosz (2017) on German demonstrative versus personal pronouns, Clem (2017) on Tswefap, and Bi and Jenks (2019) on Mandarin null versus overt pronouns—is more explicitly focused on connecting pronominal distinctions to the kinds of definiteness distinctions surveyed above. Understanding the ways that different types of pronouns map into different definite meanings is particularly important because often definite articles or demonstratives overlap considerably with pronoun paradigms.

Section 3.2 included a scale (15) of the form *exophoric>anaphoric>unique*. Pronominal distinctions also seem to make reference to this scale in that phonologically weaker pronouns, with null pronouns being the limiting case, have uses most similar to unique definiteness in a way to be made precise below, whereas phonologically stronger or overt demonstrative pronouns are typically exophoric or are used to discriminate between possible referents.[7] The tests described below might be useful for languages where a distinction between several types of pronominal reference has been identified but it is unclear what contexts license their use. As with definites, then, close inspection of the different categories of pronouns in texts is an important component of the overall descriptive picture.

Competition between pronominal forms plays an important role in explaining their distribution. For example, English *she*, *he*, and *they* allow both anaphoric and exophoric uses, whereas we have already seen that *it* and *that* contrast in this respect. The explanation is simply that demonstrative pronouns in English must be inanimate; as such they only 'compete' with *it*. So having a clear sense of the full pronominal system of a language is important in understanding why certain restrictions arise.

Now we turn to several tests for distinguishing 'weak' and 'strong' pronouns in various ways. These tests all identify contexts where a semantic contrast can be detected between a pronoun which must be semantically referential and a pronoun whose referent is licensed by uniqueness or which allows covariation.

Before we introduce specific tests, we briefly identify some general discourse-based distinctions between different classes of definites: demonstratives and phonologically stronger definites typically establish new topics, are contrastive, and are required in some syntactic contexts such as coordination or focus. In contrast, weak pronouns

[6]A point of caution is that null pronouns differ in whether they can refer to individuals or are restricted to impersonal uses (Barbosa, 2019).

[7]Note that here and below the term 'strong' and 'weak' as applied to pronouns make reference to their relative phonological weight. The relative phonological weight of pronouns is has other important grammatical consequences, such as availability in coordination and the ability to bear focus (Cardinaletti and Starke, 1999).

are often restricted to contexts where they are referring to a continuing topic, though we will see below this is not always necessarily the case.

The first test involves determining whether a pronoun can be licensed without an explicit antecedent, for example in specific situations where the referent can be contextually inferred. Only weak pronouns seem to be licensed in such cases (Patel-Grosz and Grosz, 2017). This is illustrated below for the German contrast between demonstrative pronouns and personal pronouns in (56), only the latter of which can refer to implicit antecedents, and for the Mandarin contrast between null pronouns but not overt pronouns in (57), where null pronouns are preferred in such contexts (null pronouns here and below are marked by the empty set symbol '∅').

(56) a. Wenn ich schwanger werde, werde ich {**es** / #**das**} auf jeden Fall behalten.
if I pregnant become will I it DEM on every case keep
'If I get pregnant, I will definitely keep it / #DEM(= the baby).'

b. Wenn ich ein Kind kriege, werde ich {**es** / **das**} auf jeden Fall behalten.
if I a child get will I it DEM on every case keep
'If I have a child, I will definitely keep it / DEM(= the baby).' (German, Patel-Grosz and Grosz, 2017, p. 273)

(57) a. wo ruguo huaiyun-le, jiu yiding hui liuxia **∅**/??**ta**.
I if pregnant-PERF then definitely will keep (it)
'If I get pregnant, I will definitely keep (him/her).'

b. wo ruoguo you-le haizi$_1$, jiu yiding hui liuxia **∅$_1$**/**ta**$_1$.
I if have-PERF baby then definitely will keep (it)
'If I have a baby, I will definitely keep (him/her).' (Mandarin, Bi and Jenks, 2019, p. 132)

The contrast between weak and strong pronoun mirrors the distinction between unique definites, which do not require prior mention, and anaphoric definites, which typically do.

The second test for distinguishing pronoun meanings is the case of indefinite antecedents under negation, also known as 'bathroom sentences,' due to the sentence *Either there's no bathroom in this house or it's in a funny place.* in Roberts (1989, p. 702), who attributes it to Barbara Partee. In such sentences, the indefinite in the first clause is under the scope of negation (see below on scope), so that indefinite fails to introduce a real individual who can serve as a discourse referent. Weaker pronouns can nevertheless often pick out this hypothetical individual in the second clause, while stronger pronouns cannot. For example, Bi and Jenks (2019) show that in Mandarin a null pronoun is preferred to an overt one in such a context:

(58) zhe-dong lou yaome mei-you xishoujian$_1$, yaome **∅$_1$**/#**ta**$_1$ jiu zai qiguaide
this-CLF building either not-have bathroom or (it) then in weird
difang.
place
'Either this building does not have a bathroom, or it is in a funny place.'

(59) tushuguan yaome mei-you zixishi$_1$, yaome jiu yijing youren yuding-le
library either not-have study.room or then already someone reserve-PERF
**Ø$_1$**/**#ta**$_1$.
(it)
'Either the library does not have a study room, or someone has reserved it.' (Mandarin, Bi and Jenks, 2019, ex. 131)

Bi & Jenks show the contrast between overt and null pronouns contrasts echoes a parallel contrast in Mandarin between overt demonstratives and bare nouns, where only a bare noun such as 'bathroom' could be used in place of the pronoun in (58), but a demonstrative noun phrase could not.

The 'indefinite antecedents under negation' test also reveals a contrast between two overt pronouns. In Marka-Dafing, the anaphoric definite marker *wó* has a pronominal counterpart which contrasts with the regular 3SG pronoun *á*:

(60) wá$^{!}$ábì děn-tʃémà tí tʃé míí $^{!}$bé$^{!}$é, wá$^{!}$ábì á y **=á/#=ó** dúo$^{!}$ró-só
either child-male NEG.PRES man this own or 3SG PRES -3SG/ANAPH hide-PROG.
'Either this man doesn't have a son or he's hiding him.' (Marka-Dafing, elicitation with Rassidatou Konate)

Because the anaphoric pronoun *wó* must be discourse referential, it is ruled out in this context. This test has not been widely replicated, but the distribution of pronouns in such sentences should parallel the distribution of implicit antecedents above, because the first clause fails to establish an actual discourse antecedent. As such, pronouns which require an actual entity as an antecedent seem to be impossible in such contexts.

Another test for strong versus weak pronouns is quantificational binding, which refers to sentences where the interpretation of the pronoun covaries with a quantifier occurring earlier in the sentence, as in the following English example:

(61) [Every girl]$_1$ said that **she**$_{1/2}$ was hungry.

The availability of a covarying reading for the pronoun is indicated by identical subscripts on the quantificational noun phrase and the pronoun (1, 1); having different subscripts on the quantifier and pronoun (1, 2) indicates that the pronoun must refer instead to a specific entity or discourse referent. The English sentence above does indeed show this ambiguity: it can either mean that every girl said "I am tired," or it can mean that every girl claimed about a specific girl, say, Fatimah, that she was tired.

Only some pronouns disallow bound or covarying readings. This observation was originally made by Montalbetti (1984) for Romance overt versus null subjects, only the latter of which allow bound readings. In some cases the unavailability of bound readings seems to be due to the referentiality of the pronoun in question; in other cases it seems to be due to competition with weaker forms which are preferred in binding contexts. In the examples below we see that in Tswefap (Narrow Grassfields Bantoid: Cameroon) (Clem, 2017), a 'weak' pronoun *ye* can be quantificationally bound but the 'strong' pronoun *zhig* cannot.

(62) a. [mbey weloh]$_1$ n-gop nge **yi**$_{1/2}$ a khoh.
every one TAM-say COMP 3SG FACT cough
'Everyone$_1$ said that he$_{1/2}$ coughed.'

b. [mbey weloh]$_1$ n-gop nge **zhig**$_{*1/2}$ a khoh.
every one TAM-say COMP 3SG FACT cough
'Everyone$_1$ said that he$_{*1/2}$ coughed.' (Tswefap, Clem, 2017, ex. 8)

A contrast in whether pronouns can be bound has replicated in many languages and provides valuable insight into whether a pronoun is necessarily referential in a given language, which should entail that a bound reading is unavailable. Typically, 'strong' or demonstrative pronouns which are obligatorily anaphoric or exophoric resist bound readings.

One exception to the restriction on bound readings is donkey sentences, discussed in Section 3.2.2; Bi and Jenks (2019) show that in Mandarin, stronger pronominal forms are preferred as donkey anaphora. In German, while demonstrative pronouns cannot be bound variables as above, both demonstrative pronouns and personal pronouns can serve as donkey anaphora (Patel-Grosz and Grosz, 2017).

Another context which allows discourse referential pronouns to be identified is situation-based covariation, a context which was also shown to distinguish unique and anaphoric definites (section 3.2.1). In Mandarin, null pronouns allow situation-based covariation, while overt third person pronouns do not (Bi and Jenks, 2019). This is shown for both subject and object pronouns in the following sentences:

(63) qunian, zongtong$_1$ shi minzhudang ren. jinnian, #**ta**$_1$/**∅**$_2$ shi gonghedang ren.
last.year president is democrat person this.year (he) is republican person

'Last year, the president was a democrat. This year, s/he is a republican.'

(64) zai faguo, měigeren dou xihuan zongtong$_1$. dan zai meiguo, méiren xihuan **ta**$_1$/**∅**$_2$ .
in France everyone all like president but in America nobody like (him)

'In France, everybody likes the President (=Macron). But in the US, nobody likes him (=Trump).' (Bi and Jenks, 2019, p. 132)

In sentence (63), the overt pronoun seems to refer back to the same individual president, who seems to be switching parties, rather than whoever happens to be in the role of president the following year, parallel to the sentences discussed in Section 3.2.2. In contrast, the null pronoun allows a situation-dependent covarying interpretation. In Mandarin, at least, this behavior parallels the behavior of overt and null pronouns under quantifiers in some positions, where null pronouns can be bound while overt pronouns cannot (Huang, 1991).

Another test for distinguishing weak from strong pronouns is pointing, which we saw in section 3.2.4 was a test for exophoric reference. Personal pronouns in many languages allow pointing uses, but some 'weak' pronouns do not, including Twefap *ye* or null pronouns in Mandarin:

(65) *Context:* A man walks into the room and you point to him and say:

{**zheuk** / ***yi**} a sey
3SG FACT be.tall

'He is tall.' (Tswefap, Clem, 2017, ex. 11)

(66) **ta**/#**Ø** hen congming. [pointing]
S/he very smart.
'S/he is very smart.' (Mandarin, Bi and Jenks, 2019, p. 135)

There is no assumption that pronouns in different langauges should have any restrictions on their interpretation, and in fact some pronouns seem to span all of the contexts above. For example, while demonstrative pronouns in German show restrictions in terms of much of the binding behavior above, both demonstrative pronouns and personal pronouns in German allow pointing. It is an open question exactly how many different kinds of pronouns there might be and exactly what clusters of semantic properties they might have until more languages have the kinds of contrasts described above more fully documented for their pronominal systems.

A final point, echo is that this is only there are far more pronouns than the definite pronouns surveyed here, including long-distance anaphors, which are often obligatorily bound or else receive an egophoric interpretation, or logophors. In general, logophors and long-distance pronouns tend to pattern with weak pronouns in that they tend to be bound.

**4. Indefinites** The term 'indefinite' has been used to describe a large and heterogeneous class of articles, quantifiers and pronouns (see Haspelmath 1997 for an overview). As its etymology suggests, these expressions are taken to have in common that they are not definite — a negative definition. More precisely, where the referent of a definite expression must be recoverable from context, an indefinite expression can be used to introduce a new discourse referent whose identity need not be recoverable from context, but which can be subsequently referred back to (Kamp 1981, Heim 1982). Take for instance (67), a shortened version of (4) above. In this discourse, the indefinite NP *a fifteen-foot alligator* introduces a new referent into the discourse. This discourse referent can be subsequently referred back to using the pronoun *it* (or another anaphoric expression).

(67) Did you hear what happened to me yesterday? I had an encounter with [**a** fifteen-foot alligator]$_i$. **It**$_i$ ran right at me, but I managed to escape.

While all indefinites share this core function, they differ radically in their other semantic and pragmatic features. Some of this behavior has to do competition with any definite articles the language may have, others with the core semantics of indefinites themselves — for instance whether they are quantificational or referential — and still others with additional pragmatic constraints that are at play. In this section we will cover these distinctions, providing a range of diagnostic tests.

The rest of the discussion is structured as follows. First, we give a brief survey of the common ways in which indefiniteness is marked morphosyntactically. We then provide tests to identify whether a given NP is truly indefinite, including core tests

that can be applied broadly, and those which may depend on language-specific facts. With these tests for indefiniteness established, we then turn to the distinctions found within the class of indefiniteness itself. We begin first with a discussion of how indefinites may (or may not) show scope interactions with other elements in the sentence. We then turn to licensing conditions, followed by a discussion of constraints which indefinites can place on their domain. We finish with a discussion of epistemic effects — both speaker ignorance and knowledge — that are commonly found across languages, and related pragmatics constraints, before providing a summary and discussion of how distinct properties of indefinites are known to interact.

**4.1 Markers of indefiniteness** As in many domains, the morphosyntactic realization of indefiniteness across languages is variable. Here we identify some of the common distinctions made and the morphological sources of indefinite marking.

The first distinction relevant for many languages is that between indefinite articles and pronouns. Roughly, indefinite articles are those used with an overt nominal expression, such as in (68). In contrast, pronouns, such as those in (69), do not have an overt nominal.

(68) I saw [**some/a** famous actor] the other day.

(69) I saw [**someone/something**] in the garden.

There are often morphological links between articles and pronouns, such as that found in English where the pronouns are formed by adding *one, body, thing*, etc, to an article. Some languages do not show a clear distinction between articles and pronouns. Indefinites in Tiwa, for instance, can appear with or without a nominal expression:

(70) a. Maria [**indâ-khí** kashóng-gô] pre-ga.
Maria what-INDEF dress-ACC buy-PFV
'Maria bought some dress.'

b. Maria [**indâ-khí**-gô] plaw-ga.
Maria what-INDEF-ACC forgot-PFV
'Maria forgot something.' (Tiwa)

A striking cross-linguistic pattern is similarity of the morphological source of the indefinite. Many indefinites across languages are formed from interrogative pronouns (*Wh*-words), while many others are formed from the numeral 'one'. The Tiwa examples in (70) above provide an example of this first case: one series of indefinites is formed by suffixation of *-khí* to an interrogative pronoun base. Another series, discussed below, is formed by suffixing *-pha*. Russian and Japanese indefinites are formed in a similar way, as in many other languages (Haspelmath, 1997).[8] The second case—where indefinites are derived from the numeral 'one'—is seen with indefinites in many European languages, such as with Spanish *un(a)* and *algún(a)*. It is also

[8]Note that the interrogative pronoun in these languages has often been referred to as an indeterminate pronoun, since their use in questions is one of many. See for example Shimoyama 2006 on indeterminate pronouns in Japanese.

common beyond Europe. In Ersu, a Tibeto-Burman language of Sichuan, China, the numeral tə 'one' without a classifier is interpreted as an indefinite article, but with a classifier must be read as a numeral (Zhang, 2014):[9]

(72) a. [t $s^{h}$o tə] ər=gə
dog INDEF bark=PROG
'A dog is barking.'
b. [t $s^{h}$o tə wo] ər=gə
dog one CL bark=PROG
'One dog is barking.' (Ersu, Zhang 2014:886)

Haspelmath (1997) also identifies many cases in which indefinite pronouns are built off a generic noun, such as 'person' or 'thing', as with English *something* and *somebody*.

Finally, bare nouns in many languages can also serve as indefinite NPs. This may be restricted to certain kinds of nominal expressions, such as plurals in English (73), but may hold for other bare nouns as well. For example, bare nouns in Tiwa can serve as indefinites, whether or not they have plural marking, as shown in (74). These examples come from a short text and in both cases are used to introduce the various animals that enter the scene.

(73) In Australia, Ursula was lucky enough to see [**koalas**]$_i$. They$_i$ were all sleeping at the time, but it's still pretty exciting.

(74) a. Ashôbai-lo, krom-e phána [**hâdi-raw**] arô [**makhrí-râw**] cholói phi-ga.
suddenly-FOC forest-GEN from elephant-PL and monkey-PL run AUX-PFV
'Suddenly, elephants and monkeys ran out from the forest.'
b. Peshûna, pháng-e phána [**shônggadi**] urî phi-ga.
then tree-GEN from leopard jump AUX-PFV
'Then a leopard jumped down from a tree.' (Tiwa)

An important thing to note in investigating the indefiniteness of bare nouns is that in many languages such nouns can also be interpreted as definite. For instance, sentences in Tiwa like those in (74) are often translated as definite and can be used in a context in which the NP refers back to an earlier mentioned discourse referent, as well as to introduce a new one. In such languages, indefinite and definite interpretations are not necessarily fully interchangeable. Instead, the grammatical function or position of the bare noun may affect its (in)definiteness. In Hindi and Teotitlán del Valle Zapotec, for instance, indefinite interpretations of singular bare nouns are not available in subject position (Dayal 2004, Deal and Nee 2018). Bare nouns are often also used to encode generic and kind interpretations of NPs, as discussed in Section 2. In the

[9]When just the classifier is used, Zhang reports that the NP receives a definite interpretation:

(71) [t $s^{h}$o wo] ər=gə
dog CL bark=PROG
'The dog is barking.' (Ersu, Zhang 2014:886)

remainder of this paper, we set aside bare nouns and focus instead on overtly-marked indefinites.

**4.2 Testing for indefiniteness** In this section we lay out various tests for establishing whether a given NP is indefinite. We begin with core tests which probe the basic function of all indefinites. We then turn to other potential tests, some of which depend on language-specific constructions, and others of which can only identify a subset of indefinite expressions. We then discuss some common cross-linguistic behavior of indefinites that are triggered through competition with definite expressions, applicable if the language of study has them.

**4.2.1 Core tests** The core tests for indefiniteness directly probe whether an NP can introduce a novel discourse referent that need not be recoverable from context, but which can subsequently be referred back to. The most direct test for indefiniteness, therefore, depends on texts, either naturally occurring or elicited via storyboards (and similar) or translation tasks. Essentially, if an NP can introduce a new discourse referent into a text, it is likely an indefinite. Take the English mini-discourse in (67) above, for instance. The NP *a fifteen-foot alligator* introduces a new discourse referent. Similarly the Tiwa bare noun examples in (74) above introduce new discourse referents. These textual data were elicited via a translation task, in which I provided a short English text that my consultants translated into Tiwa. This could have been easily achieved with picture- or video-based prompts as well. Naturally occurring textual data, such as traditional and personal narratives confirm that bare nouns can introduce new discourse referents.

While naturally occurring texts are an excellent way of initially identifying potential indefinite NPs, there are two limitations that caution against solely relying on them. The first applies particularly to traditional narrative texts. While to an outside researcher it may seem that a novel discourse referent is being introduced within a text, it is not necessarily the case that that discourse referent is being introduced with an indefinite. Instead, the NP may be referring to a well-known character that may be in the general story-telling context for speakers (if not the researcher). In these cases, it may be definite. For instance, it can be felicitous in English to begin a story with a definite expression, so long as the referent is familiar to speakers:

(75) A long time ago, [**the** knights of the round table] entered Camelot in order to meet with [**the** king].

This confound can be mitigated by eliciting novel texts, as described above, or by focusing on narratives of people's personal experience (as in (67)).

The second limitation has to do with the various pragmatic functions that different indefinites are associated with. If, for instance, an indefinite is used to signal speaker knowledge or speaker ignorance, it may be less likely to appear in a textual database, unless the details of the text happen to provide an appropriate discourse context for these sorts of marked indefinites. In order to mitigate this second limitation, we recommend two additional core test for indefiniteness. The first of these

is the consistency test, introduced in §3.2.1 above. This test can establish whether multiple identical NPs within a single sentence or discourse are able to introduce distinct discourse referents. If the NP is indefinite, this should be possible, but if it is definite, the two NPs would necessarily refer to the same individual. The consistency test therefore involves a sentence which should be felicitous if there are two distinct discourse referents, but infelicitous if there is only one. The test is illustrated for St'át'imcets in (76) (and its English translation). In this example, the same expression is used in a clause with incompatible predicates — a single discourse referent cannot be both in Vancouver and Mt. Currie at the same time. That the resulting sentence is felicitous shows that the NPs are indefinite.

(76) wa7 lts7a pankúph-a [**ti** swúw'h-**a**] múta7 wa7 láku7 líl'wat-a
be here Vancouver-DET [DET cougar-DET] and be there Mount.Currie-DET
[**ti** súw'h-**a**] t'it
[DET cougar-DET] also
'There is a cougar here in Vancouver and there is also a cougar there in Mt. Currie.' (St'át'imcets; Matthewson 1999: 106)

The benefit of this test in conjunction with the use of textual data is that it can be applied to any NP that is suspected of being indefinite, regardless of its frequency in texts.

There is one potential confound that must be avoided in applying the consistency test. While definite NPs will lead to contradiction and therefore infelicity, demonstratives may be used to deictically refer to two distinct referents in the immediate context (as discussed in §3.2.4 above):

(77) [**THAT** dog] is sleeping, but [**THAT** dog] is awake.

In the St'át'imcets example above, this confound is avoided since the two discourse referents in question are not in the same physical context, and an out-of-the-blue context will not provide an otherwise clear referent for each demonstrative (such as a photograph). We can therefore assume that they are not being used deictically.

An additional core test for indefiniteness has to do with the potential referent of or witness to the indefinite. As discussed in §3.2.1, many definites impose a uniqueness condition, in which there can only be one potential referent in the context that meets the NP description. Since the referent of the definite is unique in the context, it is recoverable. An indefinite, in contrast, should be felicitous even if there are multiple potential referents or witnesses in the context. Consider the context in (78), for instance. Here, the description *player* applies to many individuals in the context. Consequently, the definite is ruled out. The indefinite, however, is perfectly felicitous.

(78) Context: We are watching a football game on TV.
a. # Look! [**The** player] just intercepted that pass!
b. Look! [**A** player] just intercepted that pass!

Note that anaphoric definites (§3.2.2) can be used in contexts in which there is more than one individual that meets the description, so long as there is a linguistic antecedent to establish reference. The context in (78) rules out this confound since

there is clearly no linguistic antecedent. This context could also support a demonstrative, if the speaker is pointing or otherwise identifying the relevant player from among the others. As such, we recommend using it in conjunction with the tests described above.

**4.2.2 Other potential tests** The three tests for indefiniteness presented above probe the core function of indefinites — introducing a new discourse referent. In this section, we turn to two tests that are also helpful in distinguishing indefinites from definites, but which are more restricted in their use due to language specific factors or because they only test for certain kinds of indefinites.

The first of these involves existential constructions, which can distinguish more broadly strong vs. weak NPs (Milsark, 1974). In these constructions, only weak NPs — including indefinites — are permitted. Strong NPs — including definites — are ruled out. This contrast is shown for English in (79).

(79) a. There's [**a** fly] in my soup.
b. # There's [**the** fly] in my soup.

While this test is robust in English, many languages do not have existential constructions that exhibit this distinction.[10] For instance, the only comparable construction in Tiwa involves a verb which is can be translated as 'exist' or locative 'be'. This verb is used to convey similar existential propositions, as in (80), but can also convey locative information about a definite referent, as shown with the personal pronoun in (81).

(80) Tebul sha-w [**khum**] tong-o.
table on-LOC flower be-NEUT
'There is a flower on the table.'

(81) [**Pibúr**] nó-w tong-o.
3PL house-LOC be-NEUT
'They are in the house.'
Cf. # 'There are them in the house.' (Tiwa)

It is important therefore to establish whether there is an existential construction in the language of study that truly does make this distinction.

Another test that has been applied in identifying indefinites is sluicing. Roughly, a sluice is a partially elided clause involving a interrogative pronoun ('what', 'who') which is identified with a phrase in a preceding clause, known as the correlate. Typically sluices express uncertainty about the identity of the referent of or witness to the correlate. Crucially for our purposes, the correlate must be indefinite (or a disjunction). As shown in (82), sluicing is felicitous with an indefinite correlate, but infelicitous with a definite one.

(82) a. Anika met [**a** movie star], but I don't know who!

[10]While this distinction is not universal, it is found beyond Indo-European languages. Existential constructions in St'át'imcets, for instance, differentiate strong vs. weak determiners (Matthewson 1999), as does Mandarin (Huang, 1987).

b. # Anika met [**the** movie star], but I don't know who!

If sluicing with an NP is permitted, that is sufficient to establish whether than NP is indefinite. Note however that not every indefinite NP can serve as a correlate. In particular, some indefinites convey a sense of speaker familiarity with or knowledge of their referent or witness (see Section 4.6). In these cases, a sluice is likely to be infelicitous, just as it is with a definite. This can be clearly seen with the modified English indefinite *a certain*, which implies the speaker can identify the referent/witness:

(83) # Anika met [**a certain** movie star], but I don't know who!

**4.2.3 Competition with definites** Finally, there are a number of behaviors which many indefinites have in common by virtue of being in pragmatic competition with definite NPs. For instance, in English it is frequently infelicitous to use an indefinite when there is only one (contextually relevant) individual that meets the NP description, in other words, in immediate situation. For instance, (84) sounds generally odd and has the effect of implying that there is more than one moon. This effect is known as anti-uniqueness.

(84) # A moon is very bright tonight.

It is also infelicitous in English to use an indefinite NP to reference a previously introduced discourse referent. In the discourse in (85), for instance, the second use cannot be interpreted as the same artist that is referenced in the first sentence. That is, indefinites in English are non-anaphoric.

(85) Yesterday I met [an artist$_i$]. # [An artist$_i$] and I became friends.

While anti-uniqueness and non-anaphoric effects are common, they are not universal features of indefiniteness. This is because they arise due to pragmatic competition with definites (Heim, 1991). In English, for instance, *the* signals uniqueness and can be used anaphorically, so when an indefinite is used, it is generally interpreted as neither unique nor anaphoric. In languages which lack definite articles, in contrast, these effects disappear. St'át'imcets present a clear case of this (Matthewson, 1999). (86) shows that the indefinite determiner can be used in uniqueness contexts, and (87) shows that they can be used to reference a previously mentioned discourse referent. Because there is no definite determiner, we do not see the same anti-uniqueness and non-anaphoricity effects.

(86) ka hál'h-a [ta snéqwem-a]
OOC show-OOC [DET sun-DET]
'The sun appeared.' (St'át'imcets; Matthewson 1999: 109)

(87) a. húy'-lhkan ptakwlh, ptákwlh-min lts7a [ti smém'lhats-a] ...
going.to-1SG.SUBJ tell.story tell.story-APPL here [DET woman(DIMIN)-DET]
'I am going to tell a legend, a legend about a girl$_i$ ...'

b. ... wa7 ku7 ílal láti7 [ti smém'lhats-a]
PROG QUOT cry DEIC [DET woman(DIMIN)-DET]

> '... The girl$_i$ was crying there.' (St'át'imcets; van Eijk and Williams 1981: 19)

It can therefore be helpful to first identify whether the language under study has definite determiners. If this is the case, anti-uniqueness and/or non-anaphoricity are potential tests for indefiniteness.

Now that the basic tests for indefiniteness have been established, we turn to the ways in which indefinites vary in their semantic and pragmatic behavior, both within a single language and cross-linguistically. As alluded to above, indefinites are a heterogeneous class that, while they share a common function in introducing new discourse referents, differ broadly in their core semantics and additional pragmatic functions. We take each of these dimensions of variation in turn, beginning first with a discussion of scope and licensing, which reveals some of the most striking differences in the core semantics of indefinites. We then turn to conditions that indefinites impose on the context—specifically, on their domain—before turning to epistemic effects. In the final section, we take a broad view of the landscape, examining the ways in which these features do or do not correlate with each other.

**4.3 Scope** The scope of an indefinite has to do with its interpretation relative to other scope-taking expressions in the sentence. Take the following English sentence, for instance:

(88) Taylor wants to buy [**a** yellow hat].

This sentence is ambiguous. On one reading, Taylor has a general desire to own a certain color of hat — there is no specific yellow hat they want. On the other reading, there is one very particular yellow hat Taylor wants to buy, even if they have no desire to buy yellow hats generally. The ambiguity arises due to the interpretation of the indefinite expression relative to the intensional verb *want*. On the first reading, the indefinite is interpreted 'within the scope' of the verb: Taylor is interested in buying anything that meets the description 'yellow hat'. One the second reading, it's interpreted independently of the verb (or 'outside the scope' of the verb): there is a specific hat Taylor wants. Wide scope readings are thus often described as 'specific' readings of the indefinite, while narrow scope readings are often described as 'non-specific' (see, e.g., Haspelmath 1997). While these terms are often used in this way, they have also been used to describe the speaker's epistemic state in using an indefinite or whether the speaker intends to refer to particular individual (see Farkas 2002a, Ebert and Hinterwimmer 2013). For this reason we avoid the term 'specific' when describing scope.

Indefinites can interact scopally with a wide range of different expressions in the sentence (though not all indefinites show the same level of flexibility; see the next section). Intensional verbs such as *want* are just one kind. Other intensional verbs include *look for, hope, think* and other attitude verbs. In addition, we also find potential scope interactions with modals (*can, should, must*), negation, quantifiers like *every* and *all*, and conditionals. In all cases, the wide scope reading of the indefinite conveys that there is a single individual that serves as the witness to or referent of

the indefinite — i.e. a particular individual makes the sentence true. Narrow scope readings of indefinites, in contrast, are potentially witnessed by a wider range or individuals. An example of the scope ambiguity with a universal quantifier is shown in (89). On the narrow scope reading, there are multiple witnesses to the indefinite; each friend spoke about a separate movie. On the wide scope reading, there is a single movie.

(89) All my friends told me about [**a** new movie that came out this week].
*Narrow scope reading of* a *under* all:
Each friend told me about a different movie.
*Wide scope reading of* a *over* all:
There's a specific movie that my friends told me about.

Example (90) provides an example of this ambiguity with a conditional. The indefinite in the antecedent of the conditional can receive a narrow scope reading, on which any friend from Oakland's presence will result in box seats. It can also receive a wide scope reading in which there is only one friend from Oakland who will get us those tickets.

(90) If [**a** friend of mine from Oakland] comes, we'll get box seats at the A's-Mariners game.
*Narrow scope reading of* a *under* if:
Any friend of mine, if they're from Oakland, will allow us to get special box seats. (Perhaps there's a special deal in which any group with an Oakland resident gets those seats.)
*Wide scope reading of* a *over* if:
There's a specific friend of mine from Oakland with the relevant connections. If that friend comes, we'll get those box seats.

While indefinites like English *a* generally show scopal variability from a single surface position in the clause, a variety of factors may cause one reading to be preferred over another. These factors may be pragmatic. For instance, a particular sentence may favor one reading due to the likelihood of that reading being true or informative. There may be competition within the indefinite system itself, with one indefinite preferring a wide or narrow scope reading over others. The availability of different scope readings may also be influenced by structural position within the clause. Indefinites in higher structural positions, for instance, may necessarily receive wide scope readings. This may be the case for subjects, but may also be the case for other indefinite NPs that have scrambled or otherwise appear higher in the clause. For example, object indefinites in Hindi prefer narrow scope readings with respect to a subject quantifier in default SOV word order, but obligatorily receive a wide scope reading when they appear to the left of the subject.[11]

(91) 'Each man loves some woman.' (Hindi; Kidwai 2000:7, Bhatia and Iyer to appear:8)

[11]Bhatia and Iyer (to appear) report that a wide scope reading is available for the object indefinite in (91a), but is disprefered through competition with (91b), which is unambiguous. This is one example of how pragmatic factors may affect the availability of scope readings.

a. har aadmii [**kisii** aurat-ko] pyaar kartaa hai
each man some woman-DOM love do.IPFV be.PRES
Each man loves a different woman.
() There's a single woman who each man loves.

b. [**kisii** aurat-ko] har aadmii pyaar kartaa hai
some woman-DOM each man love do.IPFV be.PRES
Each man loves a different woman.
There's a single woman who each man loves.

**4.3.1 (In-)variable scope** While indefinites as a class can in principle show variable scope with respect to other scope taking elements, many indefinites are more restricted. In particular, languages often have indefinites which must take wide scope with respect to other elements in the clause. The indefinite determiner in the St'át'imcets examples above is one such indefinite; it cannot receive narrow scope readings (Matthewson, 1999). The same holds for the Tiwa series of *khí* indefinites Dawson 2020. (92) indicates this for a *khí* indefinite object interacting with a universal quantifier subject, but the pattern holds consistently for any other scope-taking element.

(92) Sógol mewâ-raw [**shar-khí**-gô] sêwa os-ga.
every man-PL who-INDEF-ACC greet-PFV
'Every man greeted someone.'
Every man greeted a different person.
Every man greeted a particular person.

Other obligatory wide scope indefinites include Russian *koe* indefinites (Haspelmath 1997, Kagan 2011, among others) and Sinhala *de* indefinites (Wathugala and Dawson, 2019).

While many indefinites require obligatory wide scope, other indefinites may be restricted to narrow scope in some cases. For instance, while Tiwa's *khí* indefinites must take wide scope with respect to any other scope-taking element, a separate series of *pha* indefinites must take narrow scope in conditionals:

(93) Chidî [**shar-pha** mamái] thi-gai-do, ang nó mán-o.
if who-INDEF uncle die-COND-TOP 1SG house get-NEUT
'If an uncle dies, I'll get a house.'
If any of my uncles dies, I will get a house.
There's a particular uncle who owns a house, and if he dies, I'll inherit it.

This sentence can only be felicitously used in a context in which *any* uncle's would result in the speaker getting a house; it cannot be used in a wide scope context in which there is a single uncle whose death would result in a house.

Importantly, indefinites may show different scope possibilities depending on the other scope-taking element they appear with. (93) shows that *pha* indefinites must take narrow scope in conditional antecedents. When a *pha* indefinite appears with a quantifier, it behaves similarly to Hindi *kisii* described above — it can take wide or

narrow scope depending on its surface position. When it appears with clausemate negation, in contrast, it strongly prefers a wide scope reading, as shown in (94).[12] Here, the *pha* indefinite must is interpreted outside the scope of negation: there is a particular book that Lastoi didn't buy. The narrow scope reading in which she didn't buy any books is strongly dispreferred.

(94) Lastoi [**inda-pha** lái-gô] pre-ya-m.
Lastoi what-PHA book-ACC buy-NEG-PST
'Lastoi didn't buy a book.'
Lastoi didn't buy any books.
There's a particular book that Lastoi didn't buy, (but she did buy others.)

The scope of indefinites in Ga (Kwa; Ghana) also depends on the other element they appear with. Renans (2018) reports that the indefinite *ko* can scope above or below negation, but receives narrow scope readings with respect to quantifiers. In contrast, the indefinite *kome* must scope above negation, but can receive wide or narrow scope readings with respect to quantifiers.

**4.3.2 Documenting scope** Systematically documenting the scope behavior of a given indefinite is a complex task. As we have shown, there are many factors at play, including a variety of pragmatic considerations, the position of the indefinite in the clause, and the other scope-taking element that the indefinite interacts with. While texts can provide some clues as to the scopal possibilities of a given indefinite, with a limited corpus it is unlikely that a full picture will emerge.[13] Scope judgments in an elicitation context can also be subtle and difficult to obtain, given the pragmatic variables. We strongly recommend using the semantic fieldwork methodology laid out in Matthewson 2004 to tease apart the possible readings. In this methodology, detailed disambiguating contexts are presented to the speaker (usually verbally, but visual stimuli can also be used). The speaker is then presented with the target sentence, and asked if that sentence would be acceptable in that context. This task can be repeated with different speakers and on different occasions, and with a variety of different contexts and target sentences to help mitigate pragmatic or other confounds.

An example from Matthewson's (1999) work on St'át'imcets indefinites is provided here. (95) is the sentence whose scope readings are being investigated. Note that the English translation is ambiguous – it may be that there is a specific elder whose coming will make Mary happy, or it may be that she just likes having elders around.

(95) cuz' tsa7cw kw-s Mary lh-t'íq-as [**ti** qelhmémen'-**a**].
going.to happy DET-NOM Mary HYP-arrive-3CONJ INDEF old.person-INDEF
'Mary will be happy if an elder comes.' (Matthewson 1999:90)

[12]This resistance to scoping beneath clausemate negation makes *pha* indefinites a type of positive polarity item (PPI).

[13]See Chung and Ladusaw 2003 for an example of how scope readings of Maori indefinites can be partially deduced from textual data.

The scope possibilities of the St'át'imcets indefinite were systematically tested by providing speakers one of the following disambiguating contexts, and asking if the sentence was appropriate in that context. (Note that the narrow scope context contains the clause about there being no elders to truly force a narrow scope reading of the indefinite under the conditional – it truly is a hypothetical.) The contexts can be presented in different sessions and/or the order varied between speakers to help control for unknown pragmatic variables. Using this methodology, Matthewson clearly established that the indefinite in (95) can only receive a wide scope reading.

(96) Wide scope context:
There are a bunch of elders in this community. Mary dislikes most of these elders and doesn't want them to come. There is just one elder who she wants to come.

(97) Narrow scope context:
Mary will be happy if any elders come, but that's impossible, because there are no elders in this community.

For examples of the particular kinds of contexts that can be used to tease apart indefinite scope readings in a fieldwork context, see Matthewson 1999, [xxx], and Dawson 2020. See Matthewson 2004 for a full discussion of this methodology in semantic fieldwork, which can be applied more broadly beyond investigations of scope.

While characterizing the precise scope details of an indefinite is a large and complex task, we believe it is worth investigating on at least a basic level. The scope possibilities of an indefinite are essential for understanding how an indefinite may be used in everyday contexts, as different scope interpretations often lead to very different meanings. As we have shown above, it cannot be assumed that a given indefinite will receive the same range of scope readings as more familiar indefinites such as English *a*.

Understanding the scope possibilities of an indefinite are also an essential tool for analyzing the nature of the indefinite itself. As mentioned at the outset of this section, indefinites are a heterogeneous class of expressions, united only by their core function of introducing new discourse referents. The scopal behavior of an indefinite is one of the primary diagnostics for determining the core semantic value of an indefinite–i.e., whether an indefinite is truly quantificational, or whether it is a referential expression (see Section 2). In short, if an indefinite can receive narrow scope readings, it has truly quantificational readings. In contrast, if an indefinite receives only wide scope readings, it may be a referential expression, similar to a definite (though without the need for contextual recoverability). In these cases, it would be more accurate to say that the indefinite does not take scope at all, but simply refers directly to a specific individual. Even finer-grained scope behaviors, such as intermediate scope readings, have been used to probe the nature of indefinite expressions over the last several decades (see, e.g., Fodor and Sag 1982, Abusch 1994, Reinhart 1997, Kratzer 1998, Matthewson 1999, Schwarz 2001, Brasoveanu and Farkas 2011 and Martí and Ionin 2019), and the inclusion of a broader range of cross-linguistic data is important for expanding our typology and establishing theories of indefiniteness that accurately reflect the diversity of human languages.

**4.4 Licensing** In the previous section we saw that some indefinites must be read with narrow scope when they appear with certain other scope-taking elements. A related phenomenon we find with indefinites is when an indefinite can *only* be found within the scope of another element – they are never found in positive episodic sentence that lack another scope-taking element. Such indefinites have to be licensed.

Different indefinites have different licensing conditions. A relatively common cross-linguistic licensing pattern are with indefinites that require a quantificational element higher in the clause so that the witness to the indefinite can covary with the quantifier (as in the narrow scope readings of (89) and (91)). Indefinites of this type are known as **dependent indefinites** (Farkas, 1997). An example of a dependent indefinite is given in (98). As these examples show, indefinites marked with *cîte un* are allowed when there is a higher quantifier in the clause, but ungrammatical otherwise. They always receive a narrow scope reading with respect to the licensing quantifier. While this example shows a licensing NP quantifier, dependent indefinites can also be licensed by quantificational adverbs such as *always*.

(98) a. Fiecare baiat a recitat [**cîte un** poem].
every boy has recited DEP INDEF poem
Every boy received a (different) poem.

b. * [ Cîte un student ] a plecat.
[ DEP INDEF student ] has left
Intended: 'A student has left.' (Romanian; Brasoveanu and Farkas 2011:9-10)

A common morphosyntactic form that dependent indefinites take is a reduplicated form of a non-dependent indefinite article. For instance, Hungarian has a dependent indefinite *egy-egy*, formed by reduplicating the regular indefinite *egy* (Farkas, 1997).

Other indefinites may also require licensing, but by different elements. In addition to its wide scope indefinite, St'át'imcets has an indefinite determiner *ku* which can only appear within the scope of negation and modals, and in questions Matthewson 1999. The Romanian indefinite *vreun* is only licensed in downward-entailing (DE) environments, or within the scope of an epistemic modal Falaus 2014. Other languages have indefinites that can only appear in DE environments. These indefinites are known as **negative polarity items** (NPI). An example is given here from Dutch. (99) shows the indefinite *ook maar een student* 'even one student' can be used in a negated sentence, but is ungrammatical in a positive sentence. It can be used in other DE contexts as well, such as in the antecedent of a conditional (Rullmann, 1996).

(99) a. Niemand heeft met [**ook mar een** student] overleg gepleegd.
nobody has with even one student consulted
'Nobody consulted any students.'

b. * Hij heeft met [**ook mar een** student] gesproken.
he has with even one student spoken
Intended: 'He has spoken with a student.' (Dutch; Rullmann 1996:4, 6)

Cross-linguistically NPIs are often formed with a scalar morpheme meaning 'even', though the precise licensing conditions may vary from language to language. In Tiwa,

for instance, NPIs are formed by suffixing the scalar additive particle *bo* to an indeterminate pronoun base. These NPIs, however, are only licensed under negation and in certain conditional clauses (Dawson, 2020), rather than in DE contexts more broadly.[14]

Overall, there is significant cross-linguistic variation in the licensing conditions on indefinites. Consequently, we recommend testing each indefinite in a variety of different sentence types to firmly establish whether the indefinite has any licensing requirements, including plain episodic sentences that lack any potential licensor, as well as those with quantifiers, DE contexts, a variety of modal contexts, questions and imperatives.

**4.5 Domain conditions** In §4.2.1 above, we saw that a core feature of indefinites (in contrast to definites) is that they can be used in a context in which there is more than one individual in the context that meets the indefinite description (i.e. that could serve as witness to the indefinite). For instance, the indefinite (78) above can be used when there is more than one football player. In §4.2.3, we also saw that some indefinites, such as English *a*, resist being used in contexts in which there is only one individual that could witness the indefinite. The indefinite in (84), for instance, sounds odd because there is only one contextually relevant moon. This anti-uniqueness effect arises pragmatically due to competition with a uniqueness definite – if there only one individual that meets the indefinite description in the context, then then a uniqueness definite is licensed and so is preferred due to a tendency to maximize presupposed information (Heim, 1991).

For English *a*, Heim shows that anti-uniqueness is not part of the core semantics because it can be used to introduce discourse referents whose descriptions, in all likelihood, only apply to one individual–i.e., there is only one potential witness. In (102), for instance, there is no inference that the speaker has more than one pathologically curious neighbor. Similarly, in (103) there is no inference that there is more than one 20 ft. catfish; indeed, it is unlikely that there is more than one. In these cases, anti-uniqueness effects do not arise because the individual is not recoverable from context, and so the definite article could not be used.

(102) [**A** pathologically curious neighbor of mine] broke into the attic.

(103) Robert caught [**a** 20 ft. catfish]. (Heim 1991:32)

While anti-uniqueness effects arise pragmatically for indefinites like English *a*, there are indefinites in the world's languages which do place a core semantic constraint on the size of the indefinite's domain. Specifically, there are indefinites which require

[14]Note that some – but not all – NPIs can also receive universal free choice interpretations in non-DE environments. English *any* is one such expression. (100) shows an NPI use of *anything*, licensed under negation, and receiving a narrow scope existential interpretation. (101) shows a free choice use of *anything*, which receives a universal interpretation.

(100) I didn't read [**anything**].

(101) I read [**anything** she writes]!

that there be more than individual that could serve as witness. One such indefinite is Spanish *algún* (Alonso-Ovalle and Menéndez-Benito, 2010). As (104) shows, the indefinite *algún* cannot be used when only one individual meets the indefinite description – there can only be one book that is most expensive – even when another indefinite like *un* or English *a* could be used (as shown in the translation). This example contrasts clearly with those in (103), (102) and its English translation to show that this constraint is part of the core semantics of the indefinite itself.

(104) # Juan compró [**algún** libro que resultó ser el más caro de la
Juan bought INDEF book that happened to.be the most expensive of the
librería].
bookstore
'Juan bought [**a** book that happened to be the most expensive one in the bookstore].' (Alonso-Ovalle and Menéndez-Benito 2010:16)

Alonso-Ovalle & Mendéndez-Benito describe this requirement as an **anti-singleton** domain constraint. This sort of anti-singleton indefinites is found in a variety of European languages, and more broadly, including with Tiwa's *pha* indefinites (Dawson, 2018, 2020).

A further distinction can be made among those indefinites who semantically encode that there is more than one potential witness. In particular, Alonso-Ovalle & Menéndez-Benito contrast anti-singleton indefinites like *algú*n with **domain-widening** indefinites like German *irgendein* (Kratzer and Shimoyama, 2002). Like *algú*n, *irgendein* cannot be used if the indefinite description only applies to one individual in the context. However, unlike *algún*, *irgendein* doesn't just require more than one individual in the context to be a potential witness; it requires *every* individual in the context to be a potential witness. This distinction becomes evident in certain modal contexts, such as the deontic one in (105). These sentences each express that Mary has an obligation to marry a doctor. With *algún*, the anti-singleton constraint has the effect of conveying that there is more than one suitable doctor that Mary could meet this obligation with, but perhaps there are others which are not suitable. With *irgendein*, in contrast, the domain-widening constraint has the effect of conveying that *any* doctor is a suitable candidate.

(105) 'Mary has to marry a doctor.'
(Kratzer and Shimoyama 2002:13, Alonso-Ovalle and Menéndez-Benito 2010:10)

a. María tiene que casarse con [**algún** médico].
Maria has to marry with INDEF doctor
→ there is more than one doctor that she could marry to fulfill her obligation, but there may be others that would not be allowed

b. Mary muss [**irgendeinen** Artz] heiraten.
Maria must INDEF doctor marry
→ marrying **any** doctor would fulfill her obligation

This contrast is also found with epistemic modals; anti-singleton indefinites simply require that there is more than one individual who could be the witness to the indefinite, while domain-widening indefinites require that every contextually-relevant

individual that meets the indefinite description be an epistemic possibility. As with the scope data discussed in Section 4.3 above, teasing apart an anti-singleton indefinite from a domain-widening indefinite can be tricky. Again, we recommend setting up disambiguating contexts in the manner of Matthewson 2004. For instance, the researcher could provide contexts that track on the difference in interpretation in (105). Alonso-Ovalle and Menéndez-Benito (2010: 6) provide an excellent scenario for testing the distinction with epistemic modals.

Both anti-singleton and domain-widening indefinites are associated with additional pragmatic effects, which are attributable to their domain requirements (Kratzer and Shimoyama 2002, Alonso-Ovalle and Menéndez-Benito 2010). In particular, these indefinites often convey a sense of speaker ignorance or indifference about the identity of the witness to the indefinite. (We deal with speaker ignorance effects in detail in the next section.) Domain-widening indefinites like *irgendein* have also been described as a type of free choice item, albeit of a different nature from free choice items like English *any* (see footnote 14). In particular, Chierchia 2006 draws the contrast between existential free choice items like *irgendein*, which have existential force but convey a sense of speaker ignorance or indifference, and universal free choice items like *any*, which can have universal readings.

**4.6 Epistemic distinctions** In addition to introducing a new discourse referent, many indefinites further encode extra pragmatic information about the knowledge state of the speaker with respect to the referent of or witness to the indefinite. There are two broad categories of indefinites that convey this extra information: those that convey that the speaker is ignorant with respect to the identity of the indefinite's witness/referent, and those that convey that the speaker knows their identity. Both types of indefinites have been described as epistemic indefinites, though in practice this term is typically used to describe those that convey speaker ignorance. In this section, we discuss both classes of indefinites in turn.

**4.6.1 Speaker ignorance** Many languages have indefinites that, in addition to introducing a new discourse referent, convey speaker ignorance about the identity of that discourse referent. While many indefinites, such as English *a*, are compatible with speaker ignorance, epistemic indefinites routinely and specifically convey this ignorance. A primary test for ignorance effects is in whether the indefinite can be felicitously followed up with an identification of the referent or witness. Consider, for instance, the contrast between Spanish *un* and *algún* in (106a) and (106b). The witness to the indefinite in (106a) can subsequently be felicitously identified by the speaker. The indefinite in (106b), in contrast, cannot. In using *algún*, the speaker has already conveyed that they are ignorant with respect to the witness' identity.

(106) 'Maria married a linguistics student: namely, Pedro.' (Alonso-Ovalle and Menéndez-Benito 2010:2)

a. María se casó con [**un** estudiante del departamento de lingüística]:
Maria SE married with INDEF student of.the department of linguistics
en concreto con Pedro.
namely with Pedro

b. # María se casó con [**algún** estudiante del departamento de lingüística]:
Maria SE married with INDEF student of.the department of linguistics
en concreto con Pedro.
namely with Pedro

Similarly, speakers cannot felicitously follow up an indefinite that conveys ignorance with an alternative phrase that indicates that they know the identity, such as the phrase *guess who?* (Aloni and Port, 2015). This is shown in (107) for the German epistemic indefinite *irgendein*.

(107) [**Irgendein** Student] hat angerufen. # Rat mal wer?
INDEF student has called guess PRT who
'A student has called. Guess who?' (Aloni and Port 2015:117)

A additional test for ignorance effects involves the possible responses of the addressee. Since these indefinites convey speaker ignorance, they can't be felicitously followed up with a question about their identity. This is shown in (108), for a Japanese *ka* indefinite, which conveys ignorance in unembedded contexts (Sudo 2010, Alonso-Ovalle and Shimoyama 2014). Because the speaker has already conveyed ignorance, it is odd for the addressee to ask for that information.

(108) a. John-wa kinoo [**dare-ka**-ni] atteta yo
John-TOP yesterday who-INDEF-DAT was.meeting PRT
'John was meeting somebody yesterday.'

b. # honto? aitsu dare-ni atteta?
really he who-DAT was.meeting
'Really? Who was he meeting with?' (Sudo 2010:4)

There has been a significant body of research on ignorance indefinites cross-linguistically over the past decade (see, for example, the papers in Alonso-Ovalle and Menéndez-Benito 2015). This research has shown that indefinite ignorance effects differ from each other along several dimensions, and indeed, several languages have multiple epistemic indefinites with different properties, including Sinhala (Slade, 2015) and Tiwa (Dawson, 2018).

The first dimension of variation has to do with the type of ignorance that the indefinite can encode. Above we have referred to ignorance with respect to the identity of the indefinite's witness/referent, but there are many ways one can be knowledgeable or ignorant about an individual's identity. For instance, a speaker may be able to point out the relevant individual, but may not know their name. In another case, the speaker may be able to name an individual, but be unable to provide a relevant fact about them. Aloni and Port (2015) show that epistemic indefinites can differ in their felicity conditions based on these methods of identification. For instance, while German *irgendein* can be used in contexts in which the speaker can identify the witness

by ostension (e.g. by pointing) but is unsure of their name or some salient description (e.g. their job title). In contrast, Spanish *algún* cannot be used in this context – being able to identify the witness by ostension precludes the use of this epistemic indefinite. This contrast is shown in (109).

(109) 'Look! Some professor is dancing the lambada on a table!' (Aloni & Port 2015:131)

a. Guck mal! [**Irgendein** Professor] tanzt Lambada auf dem Tisch!
look PRF INDEF professor dances lambada on the table

b. ?? Mira! [**Algún** profesor] está bailando la lambada encima de la mesa!
look INDEF professor is dancing the lambada on of the table

In contrast, both epistemic indefinites are allowed in contexts in which the speaker can identify the witness by name or some other description, but could not identify them by ostension. Differences like these can apply to different epistemic differences within a single language. Sinhala is one such language – *de* indefinites are similar to *irgendein* in that they can be used in cases where the speaker can see the witness, while *hari* indefinites pattern with *algún* (Slade, 2015).

Another way that ignorance indefinites differ from each other is in the strength and defeasability of their ignorance effects. In particular, the ignorance effects of many epistemic indefinites disappear in certain environments, usually under the scope of certain other elements. For instance, the ignorance effects that arise with Spanish *algún* and Japanese *ka* disappear under the scope of various operators, including in downward-entailing contexts and within the scope of higher quantifiers (Alonso-Ovalle and Menéndez-Benito 2010, Alonso-Ovalle and Shimoyama 2014). In (108) above we saw that it was infelicitous to ask about the identity of an unembedded *ka* indefinite, because the speaker has already conveyed ignorance. (110) shows that this is not the case for an indefinite that receives a narrow scope reading with respect to a quantifier (in this case, one in which there is a different student dancing with each professor). In this context, the speaker has not conveyed any sense of ignorance, so it is reasonable for the addressee to ask about the identities of the witnesses.

(110) a. Dono kyooju-mo [**dare-ka** gakusee-to] odotteru.
which professor-every who-INDEF student-with is.dancing
'Every professor is dancing with some student.'

b. Dare-ga dare-to odotteru no?
who-NOM who-with is.dancing Q
'Who is dancing with who?' (Japanese; Alonso-Ovalle and Shimoyama 2014:13)

In contrast, other epistemic indefinites may show strong ignorance effects in all environments. This may stem from a variety of factors. For Tiwa *khí* indefinites, ignorance effects are always present perhaps because the indefinites *cannot* scope under another operator (see §4.3.1 above). For instance, (92) above conveys that the speaker does not know which man it is that everyone met. In contrast to *khí*, which must take wide scope, the Cantonese epistemic indefinite *mzi* always shows strong

epistemic effects regardless of its scope (Lee, 2021). For example, (111) contrasts clearly with (110) above in maintaining a strong epistemic component when it receives a narrow scope reading under a quantifier – while there's a different Japanese song that's been rearranged in each case, its identity is unknown to the speaker.

(111) mai-sau batsap-nindoi coetman ge go dou hai goipin zi [**mzi bin-sau**
every-CL in.the.eighties famous GE song all be rearrange from INDEF which-CL
jatman-go].
Japanese-song
'Every famous song in the eighties is rearranged from an (unknown) Japanese song.' (Cantonese; Lee 2021:110)

These sorts of epistemic indefinites, whose ignorance effects are consistently present, are often easier to identify than those that are closer to *algún* in their behavior. For instance, Tiwa speakers frequently comment on the epistemic component of *khí* indefinites when providing translations and their intuitions about what makes *khí* different from other indefinites.

The strength and ubiquity of ignorance effects can also be seen in whether or not they can be canceled and reinforced. For instance, while *algún* and *ka* indefinites on their own convey speaker ignorance in the right context, the speaker can undo this inference by stating that they do, in fact, know who the witness is, without sounding as if they have completely contradicted themself, as shown in (112) for Japanese. (Note that this identification of the witness is marked; a simple identification in the manner of (106b) above sounds infelicitous.) Similarly, the speaker can also reinforce the inference by explicitly stating it, without sounding like they are repeating themself, as shown in (113). This dual behavior has been taken to indicate that the ignorance component of both of these indefinites is not an entailment, but instead arises pragmatically.

(112) Ken-wa [**dare-ka** gengogaku-no gakusei-to] kekkonshita. jitsuwa dare-da-ka
Ken-TOP who-INDEF linguistics-GEN student-with married in.fact who-COP-Q
shitteru
know
'Ken married a linguistics student. In fact, I know who it is.' (AO & Shimoyama 2014:14)

(113) Ken-wa [**dare-ka** gengogakka-no gakusei-to] tsukiatteiru kedo,
Ken-TOP who-INDEF linguistics.dept-GEN student-with dating but
dare-da-ka shira-nai.
who-COP-Q know-not
'Ken is dating a student in the linguistics department, but I don't know who it is.' (AO & Shimoyama 2014:14)

Among those indefinites whose ignorance component is present across the board, there is a split in whether the ignorance component can be canceled and/or reinforced. In Tiwa, the ignorance component can be, and frequently is, reinforced explicitly (Dawson, 2018, 2020)). It can also be canceled, though with more difficultly than for *algún* and *ka*. Specifically, a felicitous cancelation of the ignorance component

requires strong contextual support that includes another reason to use a *khí* indefinite, which is somewhat marked compared to other indefinites in the language. For instance, in (114), the speaker is explicitly using this marked indefinite to highlight that they will not tell the addressee who the witness is.

(114) Sonali [**shar-khí**-gô] sháre-do. Ang-do si-w shar-go, thêbo nága-do
Sonali who-INDEF-ACC flirt-IPFV 1SG-TOP know-NEUT who-ACC but 2SG.DAT-TOP
kusí-ya.
tell-NEG
'Sonali is flirting with someone. I know who it is, but I'm not going to tell you.'
(Tiwa; Dawson 2020:179)

In contrast to these other epistemic indefinites, the ignorance component of Cantonese *mzi* cannot be canceled or reinforced, as shown in (115a) and (115b) respectively. Attempted cancellation leads to a contradiction, and reinforcement sounds redundant; the speaker has already asserted that they don't know who it is. This suggests that the ignorance component of *mzi* is part if the core semantics of these indefinites (Lee, 2021).[15]

(115) Aaming tao-zi [**mzi** bin-bun syu] ...
Aaming read-PERF INDEF which-CL book
'Aaming read some book ...'
a. # ji ngo zidou hai bin-bun.
and I know be which-CL
'and I know which (book it is).'
b. # ji ngo m-zi hai bin-bun.
and I not-know be which-CL
'and I don't know which (book it is).' (Cantonese; Lee 2021:108)

In all, epistemic indefinites that convey ignorance are common cross-linguistically, but differ in the precise kinds of ignorance they convey, the contexts in which these ignorance inferences are found, and in whether they have are pragmatically derived (and therefore defeasible) or are part of the core semantics.

**4.6.2 Speaker knowledge** Many languages have indefinites that, rather than conveying speaker ignorance, convey a sense that the speaker is familiar with the referent of or witness to the indefinite. A clear case of this is seen with Russian *koe* indefinites (Haspelmath 1997, Geist and Onea 2007, Kagan 2011, among others). The use of a *koe* indefinite in (116), for instance, conveys that the speaker is able to identify the individual that came (Kagan 2011). (If the speaker cannot identify the individual, they would instead use a *to* indefinite.)

(116) [**Koe-kto**] prišel.
INDEF-who came
'Someone came.' (Kagan 2011:50)

[15]This is perhaps not surprising, as the indefinite is historically derived from *m-zi* 'not know' (Lee, 2021).

Similar effects are found with the modified English indefinite *a certain* (Hintikka 1986, Enç 1991b, Abusch and Rooth 1997, Farkas 2002b, among others). (117) comes with the inference that the speaker knows which professor released the grades early.

(117) [**A certain** professor] just accidentally released all the grades early.

A clear test for indefinites that convey speaker knowledge includes their infelicity in sluicing constructions. (118), for instance, is infelicitous because the use of a *koe* indefinite already conveyed that the speaker can identify the relevant individual. The same effect can be observed for *a certain*, as shown in §4.2.2 above.

(118) [**Koe-kto**] pozvonil, # no ja ne znaju, kto èto byl.
INDEF-who called but I NEG know who this was
'Someone called, but I don't know who it was.' (Kagan 2011:63)

Indefinites that convey knowledge often come with inference that the speaker does not wish to reveal the identity of the relevant individual. This inference arises pragmatically; typically, the speaker can cancel the inference by identifying the individual directly, as shown in (119) for Russian *koe*.

(119) [**Koe-kto**] pozvonil. Ty ne poveriš, no èto byl Dima!
INDEF-who called you NEG believe but this was Dima
'Someone called. You wouldn't believe it, but it was Dima!' (Kagan 2011:65)

As with the epistemic indefinites discussed in the previous section, indefinites that convey knowledge differ in their felicity conditions. For instance, in addition to the modified indefinite *a certain*, English also has an indefinite use of *this* (Prince 1981, Maclaran 1982, Ionin 2006). As illustrated in (120), indefinite *this* can be used to introduce new discourse referents (including in an existential construction), and further conveys that the speaker can identify the discourse referent.

(120) There is [**this** man who lives upstairs from me] who is driving me mad because he jumps rope at 2am every night. (Maclaran 1982:85)

While *a certain* and *this* both convey speaker knowledge or familiarity with the referent, this knowledge component differs in crucial ways. In particular, Ionin (2006, 2013) shows that indefinite *this* additionally requires that the speaker find the referent noteworthy. In contrast, *a certain* simply requires that the speaker be able to identify the referent with some additional property not yet mentioned (essentially being able to answer the question: *which X is it?*; Abusch and Rooth 1997, Farkas 2002b).[16]

Another point of variation is in who the knowledge effects are attributable to. Ebert et al. (2013) show that German indefinites modified by *gewiss*, for instance, strongly convey that the speaker has knowledge about the identity of the discourse referent. Those modified by *bestimmt*, in contrast, are compatible with speaker ignorance, so long as there is some other individual (e.g. the subject of the clause) that

[16]Ionin (2013) reports that the indefinite use of Russian *odin* behaves like *a certain* in this respect.

can provide that knowledge. This is illustrated in (121), which is felicitous if Peter himself knows which CD he is looking for. (In contrast, *gewiss* is infelicitous in this sentence.)

(121) Peter sucht schon seit Stunden nach [**einer bestimmten** CD] – keine Ahnung,
Peter searches already since hour after INDEF BESTIMMT CD no idea
welche genau er sucht.
which.one exactly he searches
'Peter has been looking for a certain CD for hours – I have no idea which one exactly he is looking for.' (German; Ebert et al. 2013: 38)

Indefinites that convey speaker knowledge often also take obligatory wide scope with respect to other elements (Haspelmath 1997).[17] This, combined with their epistemic effects, suggests that such indefinites are referential expressions, rather than existential quantifiers – the speaker is directly referring to a specific individual that they have in mind. Such referential indefinites are distinct from definites in that the addressee does not need to be able to recover the referent from context. While this pattern of knowledge indefinites requiring wide scope readings is very common, it does not hold universally. For example, German *bestimmt* indefinites can take narrow scope with respect to negation (Ebert et al., 2013).[18]

**4.7 Putting it all together** All languages have nominal expressions that are used to introduce new discourse referents. As we have seen in this section, however, indefinites differ along many dimensions beyond this core function. Some of these differences likely track core semantic differences, such as whether the indefinite is a truly quantification expression, or whether it is referential. Other differences have to do with the additional pragmatic functions that an indefinite performs, such as indicating speaker ignorance or knowledge with respect to the identity of the discourse referent. Documenting these differing behaviors is essential to understanding how speakers use indefinites to convey a wide range of information in their language.

In this section we have discussed several dimensions of variation. Some of these dimensions are interrelated. For instance, anti-singleton and domain-widening effects

[17]One complication is that many of these indefinites can receive functional readings, which at first glance appear to be true narrow scope readings with respect to a higher quantifier. For instance, the sentence in (122) is ambiguous between a reading on which there is a single woman whom every Englishman adores (e.g. the queen), and a reading on which there is a separate adored woman for every Englishman (e.g. his mother).

(122) Every true Englishman adores [**a certain** woman]. (Hintikka 1986:334)

Covarying readings such as these are distinct from true narrow scope readings because they are highly constrained. In particular, there must be a systematic way of selecting the woman that each man adores. In contrast, on a true narrow scope reading, this would not be the case – the sentence would be true so long as every English man adored *any* woman. This is not a reading (122) receives. For an extremely nuanced discussion of functional vs. narrow scope readings in Russian, and some excellent contexts for teasing them apart, see Martí and Ionin 2019.

[18]Since knowledge indefinites like those modified by *bestimmt* are also frequently described as 'specific' indefinites, this provides another reason to avoid the term when talking about wide scope/referential readings.

are cross-linguistically associated with speaker ignorance effects. Indefinites that convey speaker knowledge are cross-linguistically associated with obligatory wide scope. However, as more research is conducted into indefinites across languages, differing combinations of features and new behaviors are discovered, and there is no a-priori reason to assume that further semantic and pragmatic differences won't emerge.

An additional point that we have not yet addressed directly is that there is also significant variation between languages in the inventories of indefinite expressions themselves. Some languages have relatively few indefinites, while others have significantly more. While some generalizations may be made about what inventories of a certain size look like (see, for example, Haspelmath's (1997) semantic map), what kinds of indefinites a language has differs significantly across languages. For instance, while Russian has an indefinite that encodes speaker ignorance, and another that encodes speaker knowledge, Tiwa has two ignorance-conveying indefinites and none that convey knowledge. Likewise, while English, Tiwa and many other languages have flexible-scope indefinites, St'át'imcets does not – it has an obligatory wide scope indefinite, and an indefinite that must be licensed.

Finally, while we have tried to cover the major ways indefinites are known to differ both within languages and across languages, this discussion should not be taken as definitive. Instead, we hope it will be a jumping off point to identify new distinctions and to further expand our understanding of the possiblities of indefiniteness across languages.

**5. Conclusion** We have seen that while the notion of definiteness and indefiniteness are often treated as monolithic categories, in fact both are internally diverse and complex, arising due to a wide variety of contextual factors which make its reference recoverable, in the case of definites, or unrecoverable, in the case of indefinites. In both cases, the subtypes of definites and indefinite expressions show consistent semantic behavior across languages and can be identified on the basis of relatively simple semantic tests and contexts by a skilled linguist.

However, there are as many questions about these categories as answers. For example, clear cases of recognitional and salient definites discussed in Section 3.2.3 tests have not been systematically tested in the contexts used to diagnose unique and anaphoric definites. Likewise, while there are clearly identifiable broad categories of indefinites which rely on notions such as speaker knowledge or ignorance, the same semantic effects may arise through subtly different semantic and pragmatic content.

Another question that can be asked about definiteness and indefiniteness is the extent to which the pragmatic factors licensing different categories of (in)definites might be the same across the two categories. For example, the speaker knowledge condition on 'certain' indefinites in English bears a resemblance to the kinds of shared knowledge between speaker and hearer which is characteristic of salient and recognitional definites. Similarity, the uniqueness component of unique definites and the anti-singleton component of *algun* indefinites in Spanish seem to be two sides of the same coin. The goal of this enterprise, then, is finding a constrained list of possible conditions on definite and indefinite uses.

Finally, while the definite and indefinite categories recur often enough to qualify as cross-linguistically stable categories, there is no reason not to think that additional subtypes of definite or indefinite categories may exist. As such, the behaviors of definites and indefinite expressions described in this paper and the methods of identifying them simply constitute an record of what is known, in part, in order to help future linguists know when they inevitably find something new.

## References

Abbott, Barbara. 2002. Donkey demonstratives. *Natural Language Semantics* 10:285–298.

Abbott, Barbara. 2004. Definiteness and indefiniteness. *The handbook of pragmatics* 122–149.

Abusch, Dorit. 1994. The scope of indefinites. *Natural Language Semantics* 2:83–135.

Abusch, Dorit, and Mats Rooth. 1997. Epistemic NP modifiers. In *SALT VII*, ed. Aaron Lawson, 1–18. Ithaca, NY: Cornell University.

Aguilar-Guevara, Ana, Julia Pozas Loyo, and Violeta Vázquez-Rojas Maldonado, ed. 2019. *Definiteness across languages*. Berlin: Language Science Press.

Ahn, Dorothy. 2017. Definite and demonstrative descriptions: a micro-typology. *Proceedings of GLOW in Asia xi* .

Aikhenvald, Alexandra Y. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21:435–483.

Aloni, Maria, and Angelika Port. 2015. Epistemic indefinites and methods of identification. In *Epistemic indefinites: Exploring modality beyond the verbal domain*, ed. Luis Alonso-Ovalle and Paula Menéndez-Benito, 117–140. Oxford: Oxford University Press.

Alonso-Ovalle, Luis, and Paula Menéndez-Benito. 2010. Modal indefinites. *Natural Language Semantics* 18:1–31.

Alonso-Ovalle, Luis, and Paula Menéndez-Benito, ed. 2015. *Epistemic indefinites: Exploring modality beyond the verbal domain*. Oxford: Oxford University Press.

Alonso-Ovalle, Luis, and Junko Shimoyama. 2014. Expressing ignorance in the nominal domain: Japanese *Wh-ka*. In *Proceedings of the West Coast Conference on Formal Linguistics 31*, ed. Robert Santana-Labarge, 11–20. Somerville: Cascadilla Press.

Asiimwe, Allen. 2014. Definiteness and specificity in runyankore-rukiga. Doctoral Dissertation, Stellenbosch: Stellenbosch University.

Bach, Elke, Eloise Jelinek, Angelika Kratzer, , and Barbara Partee, ed. 1995. *Quantification in natural language*. Springer.

Barbosa, Pilar. 2019. 'pro' as a minimal NP: toward a unified theory of 'pro'-drop. *Linguistic Inquiry* 50:487—526.

Barlew, Jefferson. 2014. Salience, uniqueness, and the definite determiner -tè in Bulu. In *Semantics and Linguistic Theory (SALT) 24*, ed. Todd Snider, Sarah D'Antonio, and Mia Weigand, 619–639.

Barwise, Jon, and Robin Cooper. 1981. Generalized quantifiers and natural language. In *Philosophy, language, and artificial intelligence*, 241–301. Dordrecht: Springer.

Becker, Laura. 2021. *Articles in the world's languages*. Berlin: De Gruyter.

Bhatia, Sakshi, and Jyoti Iyer. to appear. Scope parallelism in Hindi-Urdu sluicing. In *Proceedings of the 7th Formal Approaches to South-Asian Languages*.

Bi, Ruyue Agnes, and Peter Jenks. 2019. Pronouns, null arguments, and ellipsis in Mandarin Chinese. In *Proceedings of Sinn und Bedeutung 23*, ed. M. Teresa Espinal, Elena Castroviejo, Manuel Leonetti, Louise McNally, and Cristina Real-Puigdollers, volume 1, 127–142. Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès).

Birner, Betty, and Gregory Ward. 1994. Uniqueness, familiarity, and the definite article in English. In *Proceedings of the twentieth annual meeting of the Berkeley Linguistics Society: General session dedicated to the contributions of Charles J. Fillmore*, 93–102. Washington D.C.: eLanguage.

Brasoveanu, Adrian, and Donka Farkas. 2011. How indefinites choose their scope. *Linguistics and Philosophy* 34:1–55.

Buba, Malami. 1997. The deictic particle CI-N in Hausa. *African Languages and Cultures* 10:29–45.

Cardinaletti, Anna, and Michal Starke. 1999. The typology of structural deficiency: On the three grammatical classes. In *Eurotype, Volume 5: Clitics in the languages of Europe*, ed. Henk van Riemsdijk, Empirical Approaches to Language Typology, 145–235. Berlin: Walter de Gruyter.

Carlson, Gregory, Rachel Sussman, Natalie Klein, and Michael Tananhaus. 2006. Weak definite noun phrases. In *Northe East Linguistic Society (NELS) 36*, ed. Chris Davis, Amy Rose Deal, and Youri Zabbal, 179–196. GLSA.

Carlson, Gregory N., and Francis Jeffry Pelletier, ed. 1995. *The generic book*. Chicago, IL: The University of Chicago Press.

Cheng, Lisa L.-S., and Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30:509–542.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339–405.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain-widening and the "logicality" of language. *Linguistic Inquiry* 37:535–590.

Christophersen, Paul. 1939. *The Articles – a study of their Theory and Use in English*. Copenhagen: Munksgaard.

Chung, Sandra, and William A. Ladusaw. 2003. *Restriction and saturation*. MIT Press.

Cisneros, Carlos. 2019. Definiteness in Cuevas Mixtec. In *Definiteness across langauges*, ed. Ana Aguilar-Guivera, Julia Pozas Loyo, and Violeta Vázquez-Rojas Maldonaldo, 39–82. Berlin: Language Sciences Press.

Clem, Emily. 2017. Two types of binding: Evidence from Tswefap pronominals. In *NELS 47: Proceedings of the Forth-Seventh Annual Meeting of the North East Linguistic Society*, ed. Andrew Lamont and Katerina Tetzloff, volume 1, 173–184.

Coppock, Elizabeth, and Stephen Wechsler. 2012. The objective conjugation in hungarian: Agreement without phi-features. *Natural Language & Linguistic Theory* 30:699–740.

Dawson, Virginia. 2018. A new kind of epistemic indefinite. In *Proceedings of Sinn und Bedeutung 22*, ed. Uli Sauerland and Stephanie Solt, volume 1 of *ZASPiL 60*, 349–366. Berlin: ZAS.

Dawson, Virginia. 2020. Existential quantification in Tiwa: Disjunction and indefinites. Doctoral Dissertation, University of California, Berkeley.

Dayal, Vaneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27:393–450.

Dayal, Veneeta. 2012. Bangla classifiers: Mediating between kinds and objects. *Rivista di Linguistica* 24:195–226.

Dayal, Veneeta, and Li Julie Jiang. 2020. The puzzle of anaphoric bare nouns in mandarin: A counterpoint to index! *Linguistic Inquiry* 1–42.

Deal, Amy Rose. 2015. Reasoning about equivalence in semantic fieldwork. In *Methodologies in semantic fieldwork*, ed. M. Ryan Bochnak and Lisa Matthewson, 157–174. Oxford University Press.

Deal, Amy Rose, and Julia Nee. 2018. Bare nouns, number, and definiteness in Teotitlán del Valle Zapotec. In *Proceedings from Sinn und Bedeutung 21*, ed. Chris Cummins, Nikolas Gisborne, Caroline Heycock, Brian Rabern, Hannah Rohde, and Robert Truswell, 317–334. Semantics Archive.

Déchaine, Rose-Marie, and Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33:409–442.

Diessel, Holger. 1999. *Demonstratives: Form, function, and grammaticalization*. Amsterdam: John Benjamins.

Diessel, Holger, and Kenny R. Coventry. 2020. Demonstratives in spatial language and social interaction: An interdisciplinary review. *Frontiers in Psychology* 11:3158. URL `https://www.frontiersin.org/article/10.3389/fpsyg.2020.555265`.

Ebert, Christian, Cornelia Ebert, and Stefan Hinterwimmer. 2013. In interpretation of the German specificity markers *bestimmt* and *gewiss*. In *Different kinds of specificity across languages*, ed. Cornelia Ebert and Stefan Hinterwimmer, 31–74. Dordrecht: Springer.

Ebert, Cornelia, and Stefan Hinterwimmer. 2013. Introduction. In *Different kinds of specificity across languages*, ed. Cornelia Ebert and Stefan Hinterwimmer, 1–10. Dordrecht: Springer.

van Eijk, J., and L. Williams. 1981. *Cúystwi malh ucwalmícwts*. Mount Currie: Ts'zil Publishing House.

Elbourne, Paul. 2013. *Definite descriptions*. Oxford: Oxford University Press.

Enç, Mürvet. 1991a. The semantics of specificity. *Linguistic inquiry* 22:1–25.

Enç, Mürvet. 1991b. The semantics of specificity. *Linguistic Inquiry* 22:1–25.

Fa□□s, Anamaria. 2014. (Partially) Free choice of alternatives. *Linguistics and Philosophy* 37:121–173.

Farkas, Donka. 1997. Dependent indefinites. In *Empirical issues in formal syntax and semantics*, ed. F. Corblin, D. Godard, and J.-M. Marandin, 243–268. Peter Lang Publishers.

Farkas, Donka. 2002a. Specificity distinctions. *Journal of Semantics* 19:213–243.

Farkas, Donka. 2002b. Varieties of indefinites. In *Semantic and linguistic theory xii*, ed. B. Jackson, 59–83. Ithaca, NY: Cornell University.

Fodor, Janet Dean, and Ivan Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5:355–398.
Geach, Peter. 1962. *Reference and Generality*. Ithaca, NY: Cornell University Press.
Geist, Ljudmila, and Edgar Onea. 2007. Specificity and implicatures. In *Proceedings of the Sixteenth Amsterdam Colloquium, December 17-19*, ed. Maria Aloni, Paul Dekker, and Floris Roelofsen, 109–114. ILLC/Department of Philosophy, University of Amsterdam.
Greenberg, Joseph. 1978. How does a language acquire gender markers? In *Universals of human language volume 3: Word structure*, ed. Joseph Greenberg, 47–82. Stanford, CA: Stanford University Press.
Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69:274–307.
Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Clarendon Press.
Hawkins, John A. 1978. *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. London: Croon Helm.
Hawkins, John A. 1991. On (in)definite articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27:405–442.
Heim, Irene. 1982. The semantics of definite and indefinite noun phrases. Doctoral Dissertation, University of Massachusetts, Amherst.
Heim, Irene. 1983. File change semantics and the familiarity theory of definiteness. In *Meaning, use, and interpretation of language*, ed. Rainer Bäuerle, C. Schwarze, and Arnim von Stechow, 164–189. Berlin: de Gruyter.
Heim, Irene. 1991. Artikel und Definitheit. In *Semantics: An international handbook of contemporary research*, ed. A. von Stenchow and D. Wunderlich. Berlin: de Gruyter.
Hengeveld, Kees. 1992. *Non-verbal predication*. De Gruyter Mouton.
Himmelmann, Nikolaus P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In *Studies in anaphora*, ed. Barbara A. Fox, 205–254. John Benjamins.
Himmelmann, Nikolaus P. 2001. Language typology and language universals. In *Articles*, ed. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, volume 1, chapter 62, 831–841. Berlin: De Gruyter.
Hintikka, Jaakko. 1986. The semantics of *a certain*. *Linguistic Inquiry* 17:331–336.
Huang, C.-T. James. 1987. Existential sentences in Chinese and (in)definiteness. In *The representation of (in)definiteness*, ed. Eric J. Reuland and Alice G. B. ter Meulen, 226–253. Cambridge, MA: MIT Press.
Huang, C.-T. James. 1991. Remarks on the status of the null object. In *Principles and parameters in comparative grammar*, ed. Robert Freidlin, 56–76. Cambridge, MA: MIT Press.
Ingason, Anton Karl. 2016. Realizing morphemes in the Icelandic noun phrase. Doctoral Dissertation, University of Pennsylvania.
Ionin, Tania. 2006. *This* is definitely specific: specificity and definiteness in article systems. *Natural Language Semantics* 14:174–234.

Ionin, Tania. 2013. Pragmatic variation among specificity markers. In *Different kinds of specificity across languages*, ed. Cornelia Ebert and Stefan Hinterwimmer, 75–103. Dordrecht: Springer.

Irani, Ava. 2019. On (in)definite expressions in American Sign Language. In *Definiteness across languages*, ed. Ana Aguilar-Guevara, Julia Pozas Loyo, and Violeta Vázquez-Rojas Maldonado, 113–151. Berlin: Language Science Press.

Jaggar, Philip John. 2001. *Hausa*. Amsterdam: Benjamins.

Jenks, Peter. 2015. Two kinds of definites in numeral classifier languages. In *Proceedings of SALT 25*, 103–124.

Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49:501–536.

Jenks, Peter, and Rassidatou Konate. to appear. Indexed definites. *Glossa* .

Kadmon, Nirit. 1990. Uniqueness. *Linguistics and Philosophy* 13:273–324.

Kagan, Olga. 2011. On speaker identifiability. *Journal of Slavic Linguistics* 19:47–84.

Kamp, Hans. 1981. A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, ed. J. A. G. Groenendijk, T. M. V. Janssen, and M. J. B. Stokhof. Amsterdam: Mathematical Centre.

Kidwai, Ayesha. 2000. *XP-adjunction in universal grammar: Scrambling and binding in HindiUrdu*. Oxford University Press.

Kouneli, Maria. 2019. The syntax of number and modification: An investigation of the Kipsigis DP. Doctoral Dissertation, New York University.

Kramer, Ruth. 2010. The Amharic definite marker and the syntax-morphology interface. *Syntax* 13:196–240.

Kratzer, Angelika. 1998. Scope or pseudoscope? Are there wide-scope indefinites? In *Events and grammar*, ed. Susan Rothstein, 163–196. Dordrecht: Kluwer Academic Publishers.

Kratzer, Angelika, and Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In *The proceedings of the third Tokyo conference on psycholinguistics*, ed. Yukio Otsu. Tokyo: Hituzi Syobo.

Lee, Tommy Tsz-Ming. 2021. Specific unknowns: A cast study of epistemic indefinites in Cantonese. *Proceedings of the Linguistic Society of America* 6:107–117.

Levinson, Stephen C. 2018. Introduction: demonstratives: patterns in diversity. In *Demonstratives in cross-linguistic perspective*, ed. Stephen C. Levinson, Sarah Cutfield, Michael J. Dunn, N. J. Enfield, and Sérgio Meira, 1–42. Cambridge University Press.

Levinson, Stephen C., Sarah Cutfield, Michael J. Dunn, N. J. Enfield, and Sérgio Meira, ed. 2018. *Introduction: demonstratives: patterns in diversity*. Cambridge University Press.

Lichtenberk, Frantisek. 1996. Patterns of anaphora in To'aba'ita narrative discourse. In *Studies in anaphora*, ed. Barbara A. Fox, 379–412. John Benjamins.

Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4:279–326.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Maclaran, Rose. 1982. The semantics and pragmatics of the English demonstratives. Doctoral Dissertation, Cornell University.
Martí, Luisa, and Tania Ionin. 2019. Wide scope indefinites in Russian: an experimental investigation. *Glossa* 4:1–31.
Matthewson, Lisa. 1999. On the interpretation of wide-scope indefinites. *Natural Language Semantics* 7:79–134.
Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70:369–415.
Matthewson, Lisa, ed. 2008. *Quantification: A cross-linguistic perspective*. Brill.
Milsark, Gary. 1974. Existential sentences in English. Doctoral Dissertation, MIT.
Montalbetti, Mario. 1984. After binding: on the interpretation of pronouns. Doctoral Dissertation, MIT.
Moroney, Mary. 2019a. Inconsistencies of the consistency test. *Proceedings of the Linguistic Society of America* 4:55–1.
Moroney, Mary. 2019b. Inconsistencies of the consistency test. *Proceedings of the Linguistic Society of America* 4:55–1–10. URL `https://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/view/4560`.
Moroney, Mary. 2021. Updating the typology of definiteness: Evidence from bare nouns in Shan. *Glossa: a journal of general linguistics* 6:1–28.
Næss, Åshild, Anna Margetts, and Yvonne Treis, ed. 2020. *Demonstratives in discourse*. Topics at the GrammarDiscourse Interface. Language Science Press.
O'Gorman, Tim. 2011. Two kinds of definiteness in Lakhota. University of Colorado at Boulder.
Partee, Barbara H. 1995. Quantificational structures and compositionality. In *Quantification in natural langauges: Volume ii*, ed. Emmon Bach, Eloise Jelinek, Angelika Kratzer, and Barbara H. Partee, 541–602. Dordrecht: Kluwer.
Patel-Grosz, Pritty, and Patrick G. Grosz. 2017. Revisiting pronominal typology. *Linguistic Inquiry* 48:259–297.
Paul, Ileana. 2009. On the presence versus absence of determiners in Malagasy. In *Determiners: Universals and variation*, ed. J. Ghomeshi, I. Paul, and M. Wiltschko, 215–245. John Benjamins.
Prince, Ellen. 1981. On the inferencing of indefinite-*this* NPs. In *Elements of discourse understanding*, ed. A.K. Joshi, B.L. Webber, and I.A. Sag, 231–250. Cambridge: Cambridge University Press.
Reinhart, Tanya. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20:335–397.
Renans, Agata. 2018. Two types of choice functional indefinites: evidence from Ga (Kwa). *Topoi* 37:405–415.
Roberts, Craige. 1989. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy* 12:683–721.
Roberts, Craige. 2004. Pronouns as definites. In *Descriptions and Beyond*, ed. Marga Reimer and Anne Bezuidenhout, 503–543. Oxford: Oxford University Press.

Rood, David S. 1942. Definiteness subcategorized in discourse: Lakhota *k'u*. In *Studies in Native American languages IV*, volume 10 of *Kansas Working Papers in Linguistics*. University of Kansas.

Royer, Justin. to appear. Decomposing definiteness: Evidence from chuj. *Canadian Journal of Linguistics* .

Rudin, Catherine. 1997. Agr-O and Bulgarian pronominal clitics. In *Formal approaches to Slavic linguistics*, volume 5, 224–252. Ann Arbor: Michigan Slavic Publications.

Rullmann, Hotze. 1996. Two types of negative polarity items. *Northeast Linguistic Society* 26.

Schrock, Terrill B. 2017. *The Ik language: Dictionary and grammar sketch*. Berlin: Language Science Press.

Schwarz, Bernhard. 2001. Two kinds of long distance indefinites. In *Proceedings of the thirteenth Amsterdam colloquium*, ed. R. van Rooy and M. Stokhof, 192–197. Amsterdam University.

Schwarz, Florian. 2009. Two Types of Definites in Natural Language. Doctoral Dissertation, University of Massachusetts, Amherst, PhD.

Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7:534–559.

Shimoyama, Junko. 2006. Indeterminate phrase quantification in Japanese. *Natural Language Semantics* 14:139–173.

Simpson, Andrew, and Priyanka Biswas. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis* 40:167–198.

Skilton, Amalia. 2019. Spatial and non-spatial deixis in cushillococha ticuna. Doctoral Dissertation, UC Berkeley.

Slade, Benjamin. 2015. Sinhala epistemic indefinites with a certain *je ne sais quoi*. In *Epistemic indefinites: Exploring modality beyond the verbal domain*, ed. Luis Alonso-Ovalle and Paula Menéndez-Benito, 82–99. Oxford: Oxford University Press.

Sudo, Yasutada. 2010. *Wh-ka* pronouns in Japanese and the semantics of indeterminate pronouns. Handout for talk at the Workshop on Epistemic Indefinites, U. of Göttingen, June 2010.

Von Fintel, Kai, and Lisa Matthewson. 2008. Universals in semantics. *The Linguistic Review* 25:139–201.

Von Heusinger, Klaus, and Jaklin Kornfilt. 2005. The case of the direct object in turkish: Semantics, syntax and morphology. *Turkic languages* 9:3–44.

Šereikaité, Milena. 2019. Strong vs. weak definties: Evidence from Lithuanian adjectives. In *Definiteness across langauges*, ed. Ana Aguilar-Guivera, Julia Pozas Loyo, and Violeta Vázquez-Rojas Maldonaldo, 83–111. Berlin: Language Sciences Press.

Wathugala, Samantha, and Virginia Dawson. 2019. In support of a choice functional analysis of Sinhala də. Ms., University of California Berkeley.

Wilkins, David. 1999. The 1999 demonstrative questionnaire:"This"and "that" in comparative perspective. In *Manual for the 1999 field season*, 1–24. Max Planck Institute for Psycholinguistics.

Wiltschko, Martina. 2012. What does it take to host a (restrictive) relative clause? *Working papers of the linguistics circle of the University of Victoria* 21:100–145.
Zhang, Niina. 2006. Representing specificity by the internal order of indefinites. *Linguistics* 44:1–21.
Zhang, Sihong. 2014. Numeral classifiers in Ersu. *Language and Linguistics* 15:883–915.

Virginia Dawson
dawsonv@wwu.edu

Peter Jenks
jenks@berkeley.edu