

Chapter 1

Weak vs. strong definite articles: Meaning and form across languages

Florian Schwarz

University of Pennsylvania

One line of recent work on definite articles has been concerned with languages that utilize different forms for definite descriptions of different types. In the first part of this paper, I discuss the semantic analysis of the underlying distinction of *weak* and *strong* definite articles as proposed in Schwarz (2009), which formalizes the contrast in terms of uniqueness (for *weak* articles) vs. anaphoricity (for *strong* articles). I also review the empirical motivation for the analysis based on German preposition-determiner contraction and its implications for related semantic phenomena. The second part of the paper surveys recent advances in documenting contrasts between definites in various other languages. One issue here will be on assessing to what extent the cross-linguistic contrasts are uniform in terms of their semantics and pragmatics, and to what extent there is variation in the relevant patterns. A second issue is to evaluate how the obvious variation in the formal realization of the contrast across languages can contribute to a more refined implementation of the contrast in meaning.

1 Introduction

Definite descriptions have played a central role in the study of meaning in natural language right from the start, going back to early work by Frege (1892), and leading to the famous debate in the philosophy of language between Russell (1905) and Strawson (1950), with continued interest in related issues (for an extensive collection, see Reimer & Bezuidenhout 2004). One central reason for this would seem to be that they offer a particularly insightful perspective on how (at least potentially) different dimensions of meaning differ from one another and interact, as well as on the role of context in interpreting linguistic utterances. Work in



linguistics has also been concerned with similar issues, specifically with regards to related questions about the interplay of contextual information and grammatical representations, in particular concerning mechanisms for quantificational co-variation, starting most prominently with Heim (1982).¹

One line of work on definite articles that has gained prominence in recent years has been concerned with languages that utilize different forms for definite descriptions of different types. While there is a fairly rich tradition in the more descriptive literature, especially on German dialects, going back at least to Heinrichs (1954), the notion that languages might have more than one type of definite article (beyond mere inflectional variations), with different semantic-pragmatic profiles, only received more wide-spread attention in the formal semantics literature in the 2000s. The present paper begins with a review of the analytical approach proposed in Schwarz (2009). It characterizes the distinction between *weak* and *strong* definite articles as in terms of uniqueness (for *weak* articles) vs. anaphoricity (for *strong* articles). The formal analysis is empirically motivated by data on German preposition-determiner contraction, and I briefly discuss the main data points in its favor, as well as its implications for related semantic phenomena.

The second part of the paper surveys recent advances in documenting contrasts between definites in various other languages. One focus here will be on assessing to what extent the cross-linguistic contrasts are uniform in terms of their semantics and pragmatics, and to what extent there is variation in the relevant patterns. A second focus is to evaluate how the obvious variation in the formal realization of the contrast across languages can contribute to a more refined implementation of the contrast in meaning, and how this relates to noun phrase structure more generally. While a fair amount of the cross-linguistic data supports the analytical contrast in terms of the weak vs. strong article distinction, there certainly is variation in definite contrasts beyond that. I briefly discuss one alternative family of proposals for capturing such variation from the literature, and also sketch some tentative analyses of additional points of variation.

Before moving on, let me issue a few caveats concerning the limitations in scope of the present inquiry. First of all, I start from the theoretical distinction I proposed in earlier work, and explore how it fares with regards to a set of cross-linguistic data that considers relevant phenomena and contrasts. This should not be taken to suggest that other theoretical approaches, beyond the ones considered here, have no role to play in the analysis of definite descriptions. Rather,

¹For a comprehensive recent proposal from the perspective of situation semantics, see Elbourne (2013).

it is simply a decision grounded in a theory-driven approach to empirical data, within which it makes sense to explore to what extent a particular analysis can deal with empirical facts. Relatedly, a core part of the proposal under consideration, as things stand, is that it makes a binary distinction. This may well turn out to be too limited, as further levels of distinction are likely to be relevant to capture all the data. Another aspect of the theoretical approach is that it takes notion(s) of definiteness developed on the basis of familiar languages such as English and German to analyze a variety of other languages. That may well come with its own pitfalls, but we have to start somewhere, and re-evaluate later to what extent those notions are suitable for spelling out the broader cross-linguistic picture. Finally, I limit my attention here to the form and meaning of definite descriptions alone, without consideration of indefinites. This, too, may be problematic in the long term, as at least some key effects in a given language may relate to the system of definite and indefinite expressions it has at its disposal. These caveats notwithstanding, I hope that the following contributes to our understanding of the typology of definiteness by evaluating a detailed formal proposal in light of a broader range of cross-linguistic data.

2 Two types of definite articles

2.1 Two semantic perspectives on definite descriptions

Broadly speaking, there are two families of approaches to analyzing definite descriptions that have been predominant in the formal literature, namely ones based on the notion of uniqueness, on the one hand, and ones based on the notion of familiarity or anaphoricity on the other hand. I provide a sketch of each of these here, following the bulk of the literature in seeing them as comprehensive proposals that aim to capture all data on definite descriptions, as is desirable for reasons of theoretical parsimony (see below for some pointers to mixed approaches in the literature).

Starting with uniqueness-based approaches, the intuitive motivation is based on examples such as the following:

- (1) Context: Speaker is standing in an office with exactly one table.
The table is covered with books.

The central idea here is that definite descriptions pick out an individual that uniquely fits the provided description. Formally speaking, the analysis is usually cast in terms of a definite description of the form *the NP* encoding that a) there is

an entity in the extension of *NP* (the existence condition) and b) that the number of such entities not exceed one (the uniqueness condition). This is at the core of both the traditions following Russell and Frege/Strawson, though they differ in the status they accord these conditions. But they agree that in the end, reference is effectively established via uniqueness (though note that they need not see the definite description itself as directly referential; Russell sees it as quantificational), so that the individual that gets talked about is precisely the one uniquely satisfying the nominal description.

For present purposes, a key point to note right away is that any analysis grounded in uniqueness faces an obvious challenge – namely that, taking (1) as our example, there are many tables in the world. The standard remedy, extensively spelled out by Neale (1990), is to appeal to a general mechanism of domain restriction, which has to be assumed independently for other kinds of noun phrases (and likely for other constructions as well). While the general idea of – and need for – such a mechanism is fairly straightforward and intuitive, its technical implementation is not, though we will not get into further detail here for reasons of space.²

One standard type of definite usage that constitutes a challenge for uniqueness-based approaches is one involving a preceding indefinite that introduces the intended referent of the definite:

- (2) a. *I got a table and an armchair delivered to my office.*
b. *The table is already covered with books.*

Crucially, and unlike (1) above, this example is perfectly compatible with there being another table in the office, which both the speaker and the addressee are aware of. The challenge for a uniqueness-based account of domain restriction then is to formulate the general purpose domain restriction machinery in such a way that the previous mention of the indefinite can bring it about that the domain only includes the newly delivered table, i.e. does not include everything in the office, even though we may very well be talking about the office as a whole in the larger conversation.

Examples like (2) constitute the core intuitive motivation for the second main approach to definite descriptions in the formal literature. It sees definites as functioning in a way rather parallel to pronouns (in a traditional view), and goes back to Christophersen (1939). The highly influential, and first fully fleshed out modern account along these lines comes from Heim (1982) (with a similar perspective

²For influential proposals, see, e.g. Westerståhl (1984), von Stechow (1994), Stanley & Szabó (2000), Elbourne (2013).

1 *Weak vs. strong definite articles: Meaning and form across languages*

offered by Kamp 1981), who proposes that definite descriptions come with an index, which has to be one that is already established, or familiar, in the discourse. The job of indefinites, in contrast, is to introduce new indices to the discourse, yielding a straightforward account of (2) as involving the establishment of an index mapped onto the newly delivered table in (2a), which is then anaphorically picked up by the definite in (2b).

As may be obvious by now, the initial example in (1) in turn constitutes a challenge for accounts based on familiarity, as there is no previous mention of the table there. The standard approach for tackling this challenge is to detach the notion of familiarity from the presence of a linguistic antecedent, e.g. by allowing entities physically present in the utterance context to count as familiar as well.³ This needs to be further extended, however, to deal with cases of so-called “global uniques”, such as *the sun* or *the pope*.

Rather than diving further into the intricacies of how each of the two accounts sketched above can deal with various challenging cases, we now turn to another perspective, which bites the bullet and admits that both analyses adequately capture how parts of natural language work. While this may seem, from an a priori perspective committed to theoretical parsimony, like admitting defeat, such an approach gains empirical motivation once languages that explicitly differentiate between different types of definite articles are considered. This is precisely the perspective put forward in Schwarz (2009), with a detailed empirical discussion of variation in contraction of definite articles and prepositions. The central argument is that certain forms (namely the contracted ones) behave exactly as expected from a uniqueness-based approach, whereas others (the non-contracted ones) exhibit the behavior we would expect from an approach that sees definites as anaphoric. To the extent that parallel patterns are found across other languages, the general empirical case for a richer theoretical inventory gets strengthened further, and one central aim of the present paper is to survey the evidence from a variety of other languages in this regard. In addition, the richer theoretical tool-box can also be put to use to deal with some of the complexities in languages without any obvious contrast between different definite articles, such as English, though that part of the story will not be pursued here, and it remains to be seen just how the English facts should be captured in light of this perspective.⁴

³For extensive discussion of the pertinent distinction between weak and strong familiarity, see Roberts (2003).

⁴For previous discussion of English data going beyond what can be captured using just one of the two approaches above, see, a.o. Birner & Ward (1994), Poesio & Vieira (1998).

2.2 Distinctions between definite articles in German and Germanic dialects

Much early descriptive work on contrasts between definite articles focused on German and Germanic dialects.⁵ The first detailed discussion of Germanic dialects with two forms for definite articles that I am aware of dates back to Heinrichs (1954), who discusses dialects of the Rhineland (see also Hartmann 1967). Other dialects for which this phenomenon has been described include the Mönchengladbach dialect (Hartmann 1982), the Cologne dialect (Himmelman 1997), Bavarian (Scheutz 1988; Schwager 2007) and Austro-Bavarian (Brugger & Prinzhorn 1996; Wiltschko 2013), Viennese (Schuster & Schikola 1984), Hessian (Schmitt 2006), and, perhaps the best documented case, the Frisian dialect of Fering (Ebert 1971a,b).⁶ A parallel phenomenon also exists in Standard German, although here the contrast is only present in particular morphological environments (Hartmann 1978; 1980; Haberland 1985; Cieschinger 2006; Puig Waldmüller 2008; Schwarz 2009). I will begin with some brief illustrations from Fering as a well-documented case with two fully distinct paradigms for definite articles, and then introduce the basic contrast in Standard German. Somewhat more subtle German data will be discussed in the following section to flesh out the nature of the contrast in meaning between the different articles.

The basic paradigm for what Ebert (1971b) calls the A-article and the D-article is presented in Table 1. The examples in (3) illustrate the contrast between the two.

Table 1: The definite article paradigms in Fering (Ebert 1971b: 159)

	m.Sg.	f.Sg	n.Sg.	Pl.
A-article	<i>a</i>	<i>at</i>	<i>at</i>	<i>a</i>
D-article	<i>di</i>	<i>det</i>	<i>det</i>	<i>dön</i>

(3) Fering (Ebert 1971b: 161)

a. *Ik skal deel tu a / *di kuupmaan.*

I must down to the_{weak} / the_{strong} grocer

‘I have to go down to the grocer.’

⁵Parts of this section are adapted from Schwarz (2013).

⁶Leu (2008) discusses related matters in Swiss German, although he focuses on syntactic issues.

1 *Weak vs. strong definite articles: Meaning and form across languages*

- b. *Oki hee an hingst keeft. *A / Di hingst haaltet.*
 Oki has a horse bought the_{weak} / the_{strong} horse limps
 ‘Oki has bought a horse. The horse limps.’

A parallel contrast can be observed in Standard German, where certain combinations of prepositions and definite determiners can, but do not have to, contract (see, among others, Hartmann 1978; Haberland 1985; Cieschinger 2006).

(4) German (Schwarz 2009: 7)

- a. *Hans ging zum Haus.*
 Hans went to_the_{weak} house
 ‘Hans went to the house.’
- b. *Hans ging zu dem Haus.*
 Hans went to the_{strong} house
 ‘Hans went to the house.’

Descriptively, the two forms seem to correspond straightforwardly to the two distinct definite articles in Fering, and I will assume in what follows that contraction reflects which article form is at play.⁷ Table 2 introduces the terminology I use to refer to the different forms, with the weak article corresponding to Ebert’s A-article and the strong one to her D-article.⁸

Table 2: Terminology for the German article forms

Form	Article type	Gloss
<i>zum</i>	weak	P_the _{weak}
<i>zu dem</i>	strong	P_the _{strong}

⁷A word of caution is in order concerning variation in contraction: some contractions are more colloquial than others, and there are corresponding differences in frequencies in written texts. My discussion focuses on prescriptively fully recognized cases, to avoid prescriptive biases against contraction, but the full range of phenomena is broader, and may even extend to differences of phonetic realization of articles in environments where contraction is not available. See Schwarz (2009: §2) for further discussion.

⁸The notions *weak* and *strong* have been used to group determiners in various other ways: Milsark (1977) used the existential construction discussed in the introduction to identify “weak” determiners, while Herburger (1997) makes yet another distinction. Finally, Carlson et al. (2006) introduce the notion of “weak definites” (with an earlier, related use by Poesio 1994), briefly discussed below. To avoid confusion, I will generally use the terms *weak article* and *strong article (definites)* in talking about the distinction introduced here.

The next section discusses the German contraction data in some detail to flesh out precisely what contrasts in meaning and use are associated with the two forms.

2.3 The contrast in meaning between weak and strong articles

The key concern for our purposes is to what extent the two different article forms differ in their meaning and conditions of use. As is the case in Fering (3), weak and strong article definites in German are not in free variation, but rather seem to be subject to different contextual constraints:

(5) German

In der Kabinettsitzung heute wird ein neuer Vorschlag vom
in the cabinet meeting today is a new proposal by_the_{weak}
{✓Kanzler / #Minister} erwartet.
chancellor / minister expected

‘In today’s cabinet meeting, a new proposal by the chancellor/minister is expected.’

The minimal contrast in availability of the weak article, based on whether the noun is *Kanzler* (‘chancellor’) or *Minister* (‘minister’) illustrates that the weak article requires uniqueness: in a given cabinet meeting, there is only one chancellor, but several ministers, thus unique reference can only be successful for the former. In contrast, the strong article does not seem to benefit similarly from contextual uniqueness:

(6) German

In der Kabinettsitzung heute wird ein neuer Vorschlag von dem
in the cabinet meeting today is a new proposal by the_{strong}
Kanzler erwartet.
chancellor expected

‘In today’s cabinet meeting, a new proposal by the chancellor is expected.’

Without further context, it is not available to refer to a minister, either, but as soon as one minister has been introduced explicitly in prior discourse, this becomes perfectly straightforward:

1 Weak vs. strong definite articles: Meaning and form across languages

(7) German

- a. *Hans hat gestern einen Minister interviewt.*
Hans has yesterday a minister interviewed
'Hans interviewed a minister yesterday.'
- b. ✓ *In der Kabinettssitzung heute wird ein neuer Vorschlag von dem Minister erwartet.*
the_{strong} minister expected
'In today's cabinet meeting, a new proposal by the minister is expected.'

Yet another example driving home the contrast between weak and strong articles is provided in (8):

(8) German (Schwarz 2009: 30)

In der New Yorker Bibliothek gibt es ein Buch über Topinambur.
in the New York library exists EXPL a book about topinambur
Neulich war ich dort und habe #im / in dem Buch nach
recently was I there and have in-the_{weak} / in the_{strong} book for
einer Antwort auf die Frage gesucht, ob man Topinambur grillen
an answer to the question searched whether one topinambur grill
kann.
can

'In the New York public library, there is a book about topinambur.
Recently, I was there and searched in the book for an answer to the
question of whether one can grill topinambur.'

Taken together, these facts suggest that uniqueness is neither necessary or sufficient for reference with the strong article. Instead, it seems to require an antecedent, here the indefinite, to refer to anaphorically. The two articles thus differ in the way they relate to their context, and they do so in a way that seems to line up rather naturally with the two main theoretical approaches to definites.

Consideration of further cases, which have been extensively discussed in the literature, extends this perspective in interesting ways. So-called bridging uses (Clark 1975; Hawkins 1978; Prince 1981) involve definites that seem to relate back to the preceding context in more indirect ways.

- (9) a. *John was driving down the street.*
b. *The steering wheel was cold.*
- (10) a. *John bought a book today.*
b. *The author is French.*

The steering wheel in (9) is of course understood as belonging to the car involved in the driving event in the first sentence. Similarly, the author in (10) is understood to be the one who authored the previously mentioned book. But how should these relations to the preceding context be seen theoretically? As it turns out, the German articles differentiate between these two standard cases in a theoretically interesting way, such that the weak article is used in the former case, but the strong article in the latter.

- (11) German (Schwarz 2009: 52–53)

- a. Part-whole relation

Der Kühlschrank war so groß, dass der Kürbis problemlos
the fridge was so big that the pumpkin without a problem
im / **#in dem** **Gemüsefach** untergebracht werden
in_the_{weak} / in the_{strong} crisper stowed be
konnte.
could

‘The fridge was so big that the pumpkin could easily be stowed *in the crisper*.’

- b. Producer relation

Das Theaterstück missfiel dem Kritiker so sehr, dass er in seiner
the play displeased the critic so much that he in his
Besprechung kein gutes Haar #am / **an dem** **Autor** ließ.
review no good hair on_the_{weak} / on the_{strong} author left
‘The play displeased the critic so much that he tore **the author** to pieces in his review.’

The first example is entirely unsurprising if we assume that the weak article requires uniqueness (plus a suitable mechanism for domain restriction, as needed for any uniqueness-based account), assuming that there is a unique crisper in the mentioned fridge. The second case is more interesting, and arguably informs just what mechanisms are at play in relating the interpretation of definites to the context. Taking the above illustrations of the role of anaphoricity for strong

1 Weak vs. strong definite articles: Meaning and form across languages

article definites seriously, the most straightforward analysis here is that the relational noun can have its relatum slot filled by an anaphoric index, which links the author directly back to the aforementioned book.

Looking beyond simple referential cases, it is well known that definites can also receive co-varying interpretations in quantificational contexts. Interestingly, both types of bridging examples (as well as ones parallel to the simple unique and anaphoric examples above) generalize to such environments:

(12) German

- a. *Jeder Student, der ein Auto parkte, brachte einen Parkschein*
every student that a car parked attached a parking-pass
am / #an dem Rückspiegel an.
on_the_{weak} / on the_{strong} rear view mirror PART
'Every student that parked a car attached a parking pass to the rearview mirror.'
- b. *Jeder, der einen Roman gekauft hat, hatte schon einmal eine*
everyone that a novel bought has had already once a
Kurzgeschichte #vom / von dem Autor gelesen.
short story by_the_{weak} / by the_{strong} author read
'Everyone that bought a novel had already once read a short story by the author.'

This is of substantial theoretical importance, as the analysis of co-variation under quantifiers is at the core of the interaction between contextual information and grammatical machinery. Thus, any analysis of the contrast between definite article forms must be rich enough to extend to a broader framework that can account for co-variation. A simple story in terms of purely pragmatic constraints on reference and contexts of use that is not tied into these more intricate aspects of grammar would thus fall short.

2.4 Sketch of the analysis in Schwarz (2009)

The core of the analysis of the two types of definites in Schwarz (2009) is that weak article definites are referential expressions (of type *e*) that presuppose that there is a unique entity meeting the description of the noun phrase (in the tradition of Frege and Strawson). In contrast, strong article definites involve an additional anaphoric component, captured by a (pronoun-like) index introduced as a syntactic argument of the strong article. The analysis is couched in a broader

framework to capture the bridging data, as well as the interplay of context and grammatical mechanisms behind co-variation in different ways for the two cases.

Starting with the weak article, the analysis assumes that a syntactically represented situation pronoun is an argument of the determiner, which provides the means for ensuring an appropriate domain restriction relative to which uniqueness holds.⁹ Semantically, the weak article denotes a function that takes a situation and a property as arguments, and returns the unique entity that has the property in that situation, if there is one (else, its denotation is undefined).

- (13) a. $[\text{DP } [the_{\text{weak}} s] \text{ NP}]$
 b. $[[the_{\text{weak}}]]^g = \lambda s_r \lambda P_{\langle e, st \rangle}. \iota x [P(x)(s_r)]$

The value of the situation pronoun is essentially determined in the same way as that of regular pronouns: it can receive its value from the assignment function, which captures the case where definites are interpreted independently of the situation relative to which the sentence as a whole is interpreted (i.e. relative to a resource situation, following the terminology of von Stechow 1994). Alternatively, it can be bound, either in such a way that it is identified with the topic situation (that the sentence as a whole is about), or by a quantificational expression, in which case the denotation of the definite as a whole co-varies with the situations quantified over.

The strong article minimally differs from the weak article in that it takes an additional individual (type e) argument, which is syntactically introduced by an index (that is semantically equivalent to a pronoun). The referent of the definite as a whole is identified with the value of this index (with the exception of bridging cases, discussed below).

- (14) a. $[\text{DP } i [[the_{\text{strong}} s] \text{ NP}]]$
 b. $[[the_{\text{strong}}]]^g = \lambda s_r \lambda P_{\langle e, st \rangle} \lambda y. \iota x [P(x)(s_r) \ \& \ x = y]$

The additional index argument of the strong article essentially introduces a familiarity constraint, as the context has to provide a value for the index via the assignment function. A preceding indefinite is one standard way for ensuring that, though other options may exist as well. While the issue of just how a referent for a strong article definite can be made familiar in a suitable way in the context deserves more in-depth exploration (also in relation to prior discussions

⁹It also accounts for the various interpretations of definites in the scope of intensional operators; see (Schwarz 2009) for detailed discussion.

of familiarity in the literature), I will limit discussion here to the former case, because it is easiest to control for in example contexts.

In addition to receiving a value contextually, the index can also be bound in various ways, rendering co-varying readings. Fundamentally, once we subscribe to the above meanings for the weak and strong articles, we are committed to allowing for both of the standard mechanisms for introducing co-variation for definites, namely via binding of the situation pronoun or of the index.¹⁰ Yet a further key consequence for interpretation in context more generally is that the specific analysis in Schwarz (2009) leaves no role to play for domain restriction via *C*-variables (basically, pronouns for predicates; see von Stechow 1994 and Stanley & Szabó 2000).

2.5 Some additional theoretical issues

While the main focus of the remainder of the paper is on cross-linguistic empirical issues, there are some further theoretical questions in relation to the analysis sketched above that should not go unmentioned (though the discussion below is hardly exhaustive in this regard). First, while the denotations in (13b) and (14b) are clearly related, and in fact largely overlap, this is not captured in any explanatory way as things stand – there simply are two lexical entries that happen to be very similar. Recent work by Grove & Hanink (2016) and Hanink (2017) proposes to address this issue by assuming just one definite article, with a denotation like the one in (13b), which can be compositionally extended to yield the strong article. In other words, the lexical variation above is instead re-analyzed as purely structural variation, all couched in a Distributed Morphology account of the contraction phenomena. This seems like a very promising avenue, though a few new questions also arise in light of it: first, given that this account is directly tied into capturing contraction, how can it be extended to languages with two full, independent paradigms for weak and strong articles (such as Fering)? Relatedly, how does this approach integrate languages where the correlate of weak article definites seems to be expressed by bare nouns? Finally, some potential evidence in favor of multiple lexical entries for different definite articles comes from Grubic (2016), who presents data suggesting a separate relational strong article variant being in play in bridging cases. Despite these further concerns, it is theoretically desirable to tie together the analysis of weak and strong articles in a more explanatory way, so reconciling these issues with a more explanatory proposal should clearly be pursued in future work.

¹⁰Given the existence of so-called donkey anaphora cases with strong article definites, the latter furthermore requires some version of dynamic binding.

Another range of rather intricate issues arises in connection with relative clauses. It has commonly been claimed in the literature that restrictive relative clauses require the strong article in their head. To the extent that this holds, it clearly requires an explanation of the interaction between the structure and meaning of the article and a relative clause structure in a position that would standardly be assumed to feature as part of its complement NP. But complicating things further, various authors have pointed out additional subtleties, potentially involving further distinctions between types of relative clauses (see, among others, Cabredo Hofherr 2013; Wiltschko 2013; Simonenko 2014). While the recent literature (including a proposal for capturing the – likely too – simple generalization about restrictive relative clauses by Grove & Hanink 2016) has contributed real advances, this area will require substantial further attention, especially cross-linguistically.

3 The weak vs. strong contrast across languages

3.1 Key empirical and theoretical questions

As we now turn to an overview of data from languages exhibiting similar phenomena, let us begin by stating the key empirical questions about the cross-linguistic data in relation to weak and strong article definites. First, we need to determine what other languages exhibit the same (or at least a highly similar) contrast in their noun phrase system. Secondly, what formal means do other languages utilize in expressing it? Finally, to what extent do we find variation in terms of its semantics/pragmatics, and how does this relate to its formal expression on the one hand and the noun phrase system of the language in question on the other?

To preview the perspective laid out below, I argue that there is quite a broad set of unrelated languages that exhibit contrasts that can arguably be modeled in a semantically uniform way, suggesting that the underlying contrast between weak and strong article definites is generally available as part of the inventory that natural languages can draw on. Within those languages, we find a wide range of formal means for encoding it. Understanding this variation in form seems crucial for a satisfactory analysis of the interplay of forms and meanings involved. In addition to this first set of languages with an essentially uniform meaning contrast, other languages seem to diverge more substantially from this pattern in that they display different types of distinctions. One possibility is that these are simply revealing yet another dimension of possible variation, that is in princi-

ple independent of the weak vs. strong contrast. Alternatively, we can consider a more gradient approach to variation, that allows languages to fall into different places of a continuum of possible differences between types of definites. Ultimately, the key theoretical questions are how many distinctions are needed to account for the range of empirical variation, what is their nature (e.g. categorical or gradient), and – if there are multiple such distinctions – how are they related? We will naturally not be able to answer all these questions conclusively, but will discuss pertinent data in relation to these issues.

With regards to variation in form, one way in which languages clearly differ is in whether they exhibit a contrast between two overt forms, or whether the contrast is between the presence and the absence of a given form (cf. the distinction between Type I and Type II splits in Ortmann 2014). The former situation clearly holds in the Germanic dialects and in Icelandic (Ingason 2016), and possibly also in Hausa and Lakhota (for discussion and references, see Schwarz 2013). The latter situation seems to hold in Akan (Arkoh & Matthewson 2013), Korean (Cho 2016; Ahn 2016), Mauritian Creole (Wespel 2008), Czech (Šimík 2015), Thai and Mandarin (Jenks 2015), Upper Silesian (Ortmann 2014), Upper Sorbian (Ortmann 2014), Ngamo (Grubic 2016), American Sign Language (Irani & Schwarz 2016) and Lithuanian (Šereikaitė 2016).

The following sections provide illustrative pairs of examples from a fair number of these languages, selected to highlight cases where the contrast has been studied in some detail. The core phenomenon I focus on is bridging, as this is both in many ways the most subtle and perhaps most surprising aspect of the article contrast, since the data themselves in no way intuitively impose what analysis of definites would be the most obvious candidate. But note that at least generally speaking, parallel effects systematically occur for more standard anaphoric and unique definite uses in all these cases, so the data discussed here for illustration should not be taken to suggest that the relevant distinction is only made for the bridging cases.¹¹

3.2 Illustrations of weak and strong article definites across languages

The first illustration comes from Akan. Arkoh & Matthewson (2013) discuss data parallel to that considered in Schwarz (2009), with a contrast between bare noun phrases, as in (15a), which presumably is a case of bridging involving situational

¹¹A caveat before diving into the cross-linguistic data: not all of the languages discussed below have been investigated at the same level of empirical depth, and there thus may be more variation than apparent here. But I tried to only include relatively well-documented cases that so far have essentially yielded complete overlap with the German contrast.

uniqueness, and the familiar form *nɔ* in (15b), which they argue to be a case of anaphoric bridging.¹²

(15) Akan (Arkoh & Matthewson 2013: 14–15)

a. Weak

Yè-hú-ù dàn dádáw bí wò èkùràsí hó ñkyénsidán
 1PL.SBJ-see-PAST building old INDEF at village there roof
 (#*nɔ* / #*bi*) é-hódwòw
 DEF / INDEF PERF-WORN-out

‘We saw an old building in the village; (#**the** / #**a (certain)**) roof was worn out.’

b. Strong

Ásáw nɔ yé-è ðhín nɔ fêw árá mà ð-kyé-è
 dance DEF do-PAST chief DEF beautiful just COMP 3SG.SBJ-give-PAST
 ðkyìréfɔ́ *nɔ* àdzí
 trainer FAM thing

‘The dance was so beautiful that the chief gave **the trainer** a gift.’

Similarly, Mauritian Creole, discussed by Wespel (2008), distinguishes between a null form (16a) and one clearly derived from the French definite article *la*, but which seems to be restricted to uses parallel to the strong article, as illustrated by the anaphoric ‘book-author’ bridging case in (16b).

(16) Mauritian Creole (Wespel 2008: 155–156; source: O.M.2.8, O.M.22)

a. Weak

*Mo fin visite enn lavil dan provins. **Lameri** ti pli ot ki*
 I ACC visit one village in province town-hall PST more high than
legliz.
 church

‘I visited a village in the province. **The town hall** was higher than **the church.**’

b. Strong

*Li fin kontan liv la ek aster li envi zwen **loter** la.*
 she PST love book DEF and now she want meet author DEF

‘She was fond of the book and now she wants to meet **the author.**’

¹²For recent work offering a different perspective, which disagrees with the familiarity-based analysis by Arkoh & Matthewson (2013), see Bombi-Ferrer (2017).

1 Weak vs. strong definite articles: Meaning and form across languages

American Sign Language features an expression resembling pointing within the signing space, which has been much discussed in the recent literature with regards to its pronominal uses (Schlenker 2017). However, it also serves the role of a strong definite article, as illustrated by its obligatory occurrence in anaphoric bridging in (17b).¹³ In contrast, cases involving situational uniqueness bridging, as in (17a), are incompatible with this form.

- (17) American Sign Language (Irani 2016)
- a. Weak
IX_a CAR, POLICE STOPPED WHY (#IX_a) MIRROR BROKEN.
'The car was stopped by the police because the mirror was broken.'
 - b. Strong
JOHN BUY IX_a BOOK. #(IX_a) AUTHOR FROM FRANCE.
'John bought a book. The author is from France.'

In yet another similar vein, recent discussion of Korean suggests that what had traditionally been considered a demonstrative – *ku* – seems to function as a familiar definite marker, while uniqueness based definites are expressed with bare noun phrases.¹⁴

- (18) Korean (Cho 2016: 6)
- a. Weak
Gyeolhonski-e gatda. Sinbu-ga / #ku sinbu-ga paransek-ul
wedding-to went bride-NOM / that bride-NOM blue-ACC
ipeotda.
wore
'(I) went to a wedding. The bride / #that bride wore blue.'
 - b. Strong
Jonathan-un eojebam-e sesigan dokseorul haetda. ku
Jonathan-TOP yesterday night-at three hours reading did. ku
soseolchayk-i / #soseolchayk-i jaemi-itdago saengakhaetda.
novel-NOM / novel-NOM interesting thought.
'Jonathan read for three hours last night. (He) found the novel interesting'

¹³Interestingly, this same form can also be used to introduce new discourse referents, as can be seen in the first sentence of (17b); see Irani (2019) [in this volume] for a fuller analysis.

¹⁴See Ahn (2017) for a recent proposal that Korean actually makes a three-way split, further extending the typological picture.

A final case (at least as far as the present discussion is concerned) of a language that has been argued to feature an overt form, namely a specific classifier construction, that parallels strong article definites, vs. bare nouns to express weak article definites, is that of Thai.

(19) Thai (Jenks 2015: 109)

a. Weak

rót khan nán thùuk tamrùat sàkàt phrǝʔ mâj.dâj tit
 car CLF that ADV.PAS police intercept because NEG attach
satikǝ wâj thii thábian (#baj nán).
 sticker keep at license CLF that

‘That car was stopped by police because there was no sticker on the license.’

b. Strong

?ɔl khít wâa klɔn bòt nán prǝʔ mâak, mǝɛ-wâa kháw cà
 Paul thinks COMP poem CLF that melodious very, although 3SG IRR
mâj chǝɔp náktǝɛŋklɔ # (khon nán).
 NEG like poet CLF that

‘Paul thinks that poem is beautiful, though he doesn’t really like the poet.’

A rather different instantiation of the weak vs. strong article contrast can be found in Icelandic. While the definite article generally appears as a suffix on the head noun, this suffixation is blocked by a certain class of evaluative adjectives. Ingason (2016) shows that the free form *hinum*, which had previously been considered as archaic, can occur in such cases in the modern standard, but only if we are dealing with a weak article definite. Strong article definites in such circumstances can only be expressed by the demonstrative *þessum*.

(20) Icelandic (Ingason 2016: 108, 131)

a. Weak

Context: The speaker is annoyed that she always loses. There is only one winner per round.

Alltaf eftir hverja umferð eru spilin gefin aftur af
 always after each round are cards.the given again by
 [DP **hinum** óþolandi sigurvegara].

HI-the_{weak} intolerable_{evaluative} winner

‘Always after each round, the cards are dealt again by the intolerable winner.’

1 Weak vs. strong definite articles: Meaning and form across languages

b. Strong

Previous discourse: Mary talked to a writer and a terrible politician.
She got no interesting answers from...

...þessum / #hinum hræðilega stjórnmálamanni.

...this / HI-the_{weak} terrible_{evaluative} politician

Another case where adjectives crucially feature in the expression of the weak vs. strong contrast, though in a different way, is Lithuanian (Šereikaitė 2019 [in this volume]). It exhibits a definite suffix that appears on adjectives, but only when they are of the strong article definite variety. In cases of uniqueness-based definites, the adjective will form a noun phrase with the noun without this suffix. Interestingly, such “bare” forms also have indefinite uses. Furthermore, the suffix has a much wider distribution, and can also appear on demonstratives and pronouns, among others. This wider distribution, as well as more intricate variations in the range of uses involving kind reference, deserve much more detailed attention, but at this point it seems safe to say that at least part of the contrast between bare and definite-suffixed forms seems to track the weak vs. strong article definite contrast.

(21) Lithuanian (Šereikaitė 2019 [in this volume])

a. Weak

Praėjus dviem savaitėm po rinkimų, prezidentas turi teisę atleisti

Passed two weeks after elections president has right fire

naują / #naują-jį ministrą pirmininką tik išskirtiniais atvejais.

new / new-DEF minister prime only exceptional cases

‘Two weeks after the election, the president has a right to fire **the new prime minister** only in exceptional cases.’

b. Strong

Knyga “Lietus” sulaukė neįtikėtino populiarumo, nepaisant to, kad

Book ‘Rain’ received incredible popularity despite that

talentingas-is / #talentingas rašytojas nusprendė likti

talented-DEF_{strong} / talented_{weak} writer decided remain

anonimas.

anonymous

‘The book ‘Rain’ became incredibly popular despite the fact that **the talented writer** decided to remain anonymous.’

While this overview can only be cursory, given space constraints, the relatively minimal pairs of examples from this range of largely unrelated languages

should illustrate that key phenomena concerning the weak vs. strong-article definite contrast are mirrored by formal distinctions between different types of definite noun phrases cross-linguistically. There are two key questions, both from a theoretical perspective and for pursuit in future research on definites across languages: a) how does the formal expression of the contrast vary across languages and how does this variation relate to the core meaning contrast? b) to what extent is the contrast the same across languages, and to what extent, and in what form, do we find variation in this regard. I turn to some – necessarily preliminary – considerations in the following section.

4 Variation in form and meaning

4.1 Variation in form

Starting with variation in the form of how the contrast between weak and strong article definites is expressed, an initial generalization, from the perspective of the analysis of Schwarz (2009), seems to be that a ‘more’ in meaning is generally reflected in a ‘more’ in form: the weak article definites in German and related dialects all involve morpho-phonologically reduced forms, e.g. contraction in Standard German. In the Germanic dialects with two full article paradigms, weak article forms also seem to be less complex than strong article ones. And in many languages, of course, this situation descriptively holds in the extreme, as weak article definites are expressed with bare noun phrases.

Two particularly interesting cases with regards to the formal realization of the contrast are Icelandic and Lithuanian. In Icelandic, the same nominal suffix is used to express both types of definites in most contexts. Only when, in the analysis of Ingason (2016), suffixation is blocked by evaluative adjectives do we find a distinction, such that an otherwise archaic free-form article is used for weak article definites. While at first sight, this seems perhaps at least in one sense more complex than the default configuration, strong article definites cannot be realized by the default form in that case either, but instead call for a demonstrative (which *is* more complex).

Turning to Lithuanian, the perhaps most notable point is that the explicit indication of definiteness occurs neither on the noun itself or at the level of a (potential) D-head, but rather in the form of a suffix on adjectives between these two. The formal relation between this suffix and a potential null D-head of course constitutes one key question in this regard, and there seem to be arguments in favor of a DP-layer for both cases, contrary to what has been said about, e.g.

Serbo-Croatian, where the formal realization otherwise seems somewhat similar (Šereikaitė 2016). In addition, it bears repeating that the same suffixal form that we find on adjectives can also appear in various other places, most relevantly pronouns and demonstratives. While in principle, the effect there does not seem to be dissimilar, the details are not obvious and require much more extensive exploration.

Returning to the more general issue of meaning and form, the apparent generalization about the formal realization of the distinction should be taken seriously and relates to key choice points in the semantic analysis of the article contrast: if we want to capture the relationship between both the forms and meanings involved in such a way that one is in some way derived from, or an extension of, the other, then this would call for broader proposals of the sort put forth by Grove & Hanink (2016) and Hanink (2017), briefly discussed above, which extend to cases of languages with two full article paradigms. On the other hand, if we assume two distinct lexical entries for weak and strong articles, then the generalization about the forms involved would have to be explained in another way, e.g. from the perspective of historical development, which could see the morpho-phonologically less complex forms as more grammaticalized or bleached, perhaps in parallel to the relation between demonstratives and definite articles more generally (Lyons 1999).

The fact that many languages use bare noun phrases for the weak article also relates to this question, of course, as well as to key issues in DP-syntax. In particular, the question arises of whether or not a determiner-level is present in these noun phrases in the first place, and if so, why it is the weak article meaning that can standardly be realized as phonologically null. Alternatively, a common move is to assume that purely semantic type-shifters can do the job of (both definite and indefinite) articles when overt forms are lacking (Partee 1986; Chierchia 1998; Dayal 2004). This then raises questions about the interplay between the determiner-inventory in the relevant languages and the constraints for the applications of such type-shifters. Furthermore, since the null-hypothesis for such type-shifters clearly would be that their effect is universal across languages, any variation in the interpretive options of bare noun phrases that cannot be accounted for in terms of the determiner system of the language in question, e.g. in terms of blocking effects from available overt forms, would seem to support the notion that distinct lexical determiners with the same phonologically null form can in principle be available, in contrast to what is commonly argued by proposals based on type-shifters (for recent discussion, see Dayal 2016).

Of particular importance in this regard is the potential case of languages which exhibit a genuine ambiguity between definite and indefinite interpretations for bare noun phrases. Initial evidence in relevant discussion of, e.g. Akan (Arkoh & Matthewson 2013), Lithuanian (Šereikaitė 2019 [in this volume]), and ASL (Kouli-dobrova 2012; Irani 2019 [in this volume]) suggests that this is a possibility, contra the type-shifter based proposal by Dayal (2016), but further scrutiny is needed, both empirically and in terms of integrating the article-contrast issues into the broader theoretical picture.¹⁵

4.2 Variation in meaning

While in the data so far the semantic contrast arguably can be seen as entirely uniform, it is undeniable that there is some degree of variation in this regard as well. Some of it consists of fairly detailed aspects, including what forms are used in certain cases where the contextual constraints for anaphoric uses or situational uniqueness are met, and in some cases additional distinctions involving other features may be at play as well. Generally speaking, these cases are consistent with the semantic analysis of the contrast laid out above, but involve differences in what form winds up being preferred given a certain type of context. But there also seems to be more substantial variation, which may require reconsidering the broader theoretical set of options. Some illustrations of the former cases are provided in the remainder of this section, while I turn to the latter in the next section.

One point of more subtle variation concerns anaphoric usage in longer narrative texts. A central character of a story (e.g. a fisherman, as in the Fering story considered by Ebert 1971b) may be introduced with an indefinite, and then initially picked back up by a strong article definite. But as the central role of the character becomes clear in the narrative, one may then switch to using weak article definites for it. In contrast, according to intuitions reported by Anton Ingason (p.c.), Icelandic would keep using the form corresponding to the strong article definite in this situation. But while the conditions for anaphoric uses are met, the central role of the character in question may also suffice to provide contextual restriction to ensure uniqueness of that entity.

Another point of variation concerns contexts involving entities which are both unique and familiar (at least in a weak sense) in the broader non-linguistic con-

¹⁵One important question in this discussion is what counts as an “article-less” language for the purposes of generalizations made by such proposals: where do languages which express weak article definites with bare noun phrases, but have an explicit determiner form for strong article definites, fall?

1 Weak vs. strong definite articles: Meaning and form across languages

text, e.g. with regards to a family dog. Akan and German seem to differ here, in that the former chooses to use the overt strong article, whereas German prefers the weak article form.¹⁶

(22) Context: You and your spouse own one dog. While your spouse is away, someone breaks into your house and you are telling them about it on the phone. You say:

a. German (Arkoh & Matthewson 2013: 19)

Der Einbrecher ist zum Glück vom / #von dem Hund
the burglar is to_the_{weak} luck by_the_{weak} / by the_{strong} dog
verjagt worden.
chased been

‘Luckily, the burglar was chased away by the dog.’

b. Akan (Arkoh & Matthewson 2013: 19)

Òwifó nɔ́, bɔ́dóm nɔ́ kà-á nɔ́-dɔ́ árá má
thief DEF dog DEF follow-PAST 3SG-OBJ-ON just so
ò-gúán-ìì.
3SG.SBJ-RUN-PAST

‘The thief, the dog chased away.’

But as before, the fact that conditions for situational uniqueness are met and an anaphoric form is used is not incompatible with the formal analysis. All that is required for a strong article definite is that its index receives a value from the assignment function. When an entity such as a family dog is familiar in a context, that may suffice to establish that, parallel to how personal pronouns can be used in similar situations, e.g. by parents who have a single boy who can be referred to as *he* without any recent prior mention. But nonetheless, the question, of course, needs to be addressed just why a language like Akan should differ precisely in that regard from other languages. One possibility is that the availability of indefinite uses of plays a role here; this will need to be tested with regards to other languages with similar properties.

Contexts of situational uniqueness bridging also seem to exhibit some variation. For example, Wespel (2008) cites Amern data from Heinrichs (1954), showing that the strong article is used in the following example for the noun phrase headed by *altars*, even though it is clearly part of the aforementioned church.

¹⁶Mauritian Creole may be similar to Akan in this regard; see Wespel (2008: 189–190).

(23) Amern (Heinrichs 1954: 99)

Vör worən en dā nāldər kerək on wolən os äns di altöors
we were in DEF of-N church and wanted us once DEF.PL_{strong} altars
bekikā.
look-at

‘We were in the church of Waldniel and wanted to have a look at the altars.’

The extent to which this is compatible with the formal analysis at least in part depends on the properties of the nouns in question, in particular with regards to the possibility of them receiving a relational meaning, as relational nouns in principle will open up to anaphoric bridging with the strong article, parallel to the book-author cases considered above. Interestingly, other languages have been argued to exhibit inter-speaker variation precisely in this regard: Ortmann (2014) reports data from Upper Sorbian, which seems to at least in part reflect generational variation such that, for some speakers, the strong article *tón* is not obligatory in cases like the following, while it is obligatory across the board in cases parallel to the book-author examples. Additionally, Ortmann reports parallel judgment patterns in Upper Silesian to be extremely hard to ascertain empirically.

Yet another dimension of potential minor variation involves additional distinctions. In particular, Ahn (2016) reports a 3-way split in Korean, with an additional form specialized for genuinely deictic uses (which are commonly available for strong article forms in other languages as well).

In sum, there is clear evidence of what can be considered fairly minor variation in the article contrast across languages, which in principle is consistent with the semantic characterization provided, but calls for further explanation of why languages should make different pragmatic choices about which article to use in a given type of context. Additionally, further and more fine-grained distinctions extending beyond the weak-strong contrast seem to exist as well. While much more needs to be explored, this data at least in principle seems to be amenable to explanation within the general approach outlined above.

5 Beyond weak vs. strong

5.1 Different semantic contrasts

In addition to what we saw in the previous section, there are other languages that seem to diverge in more substantial ways in the way that they exhibit a contrast between different types of definite articles. For example, while Haitian Creole is superficially similar to Mauritian Creole, and both have French as their main source language, the contrast between definite noun phrases marked with *la* (derived from the French definite article, as in Mauritian Creole) and bare ones seems different from what we have seen before.¹⁷ First, parallel to the Amern data above, there seems to be no contrast between different types of bridging, and both situational and anaphoric bridging use the overt form (here realized as *la* or *a*):

(24) Haitian Creole (Wespel 2008: 114; source: E.F.32, E.F.36.9)

a. Weak article definite context

Yè, mwen viste yon vil provens. Meri a pi wo
yesterday I visit one town province town-hall DEF more high
ke legliz la.
than church DEF

‘Yesterday I visited a town in the province. **The town hall** was higher than **the church**.’

b. Strong article definite context

Eli te renmen liv la, e kounye a li vle rankontre otè
Eli PST love book DEF and now DEF she want meet author
a.
DEF

‘Eli loved the book, and now she wants to meet **the author**.’

Similarly, larger or immediate situation uses (in the terminology of Hawkins 1978), which in other languages call for the weak article or equivalent, also generally call for the overt form. The bare form is only used for what Wespel calls complete functional descriptions, i.e. cases where the head noun denotes a function and its relatum argument is explicitly introduced, as in (25), which, as Wespel spells out in some detail, does not involve a possessive construction of any sort.

¹⁷Potential other candidate languages fitting this category include Bangla (Simpson & Biswas 2016) and Jinyun (Simpson 2017), though further research is needed to compare these various cases in more detail.

- (25) Haitian Creole (Wespel 2008: 98)

papa Mari
‘the father of Mary’

This situation seems very much at odds with the weak vs. strong article contrast as spelled out above. To begin with, global uniques (such as *the sun*) are core cases for the analysis in Schwarz (2009). The split between these and “complete functional descriptions” is also rather puzzling from that perspective. One sensible reaction might be to take this to reflect a fundamentally different contrast, and I will explore some potential avenues for such a move below. But even if this were successful, it would leave us with vexing questions about how this state of affairs came about, especially given the fairly minimal pair of two French-based creoles that both retain a form based on French *la*, but use it in apparently very different ways.¹⁸

Turning to potential directions for alternative characterizations of the Haitian Creole contrast, some rather suggestive examples are discussed by Wespel (2008). In particular, the presence or absence of *la* seems to relate to the introduction of the domain of *only* (and parallel effects exist for superlatives). In particular, when the domain of *only* is explicitly restricted by a post-nominal prepositional phrase, such as ‘in his family’, then no *la* (or allomorph) appears on the noun phrase associated with *only* (26a). In contrast, when this prepositional phrase is used as a framing adverbial, and not in the scope of *only*, then the overt article form does appear (26b).

- (26) Haitian Creole (Wespel 2008: 118–119; source: E.F.76.20.a, E.F.76.20.b)

- a. *Pyé se sèl gason nan fanmi li.*
P COP only boy in family his
‘Peter is **the only boy** in his family.’
- b. *Fanmi sa a, se yon gwo fami, men Pyé se sèl gason an.*
family DEM DEF COP INDF big family but P COP only boy DEF
‘This family is big, but Peter is **the only boy**.’

Given this suggestive data, one potential avenue to explore, building on the proposal by Wespel (2008) that *la* indicates the use of a “resource situation variable”, is that it is the overt realization of a situation pronoun in the sense of

¹⁸Another interesting potential consequence of such a move, which I am not able to explore here in detail, is that this would seem like another case of genuine variation in the type of definiteness involved with bare noun phrases, which would come as somewhat surprising for type-shifting based accounts of such noun phrases, again under the assumption that what type-shifters can do is universal.

Percus (2000). Formally, a candidate requirement introduced by this particular type of situation pronoun could be that it is not identical to the topic situation relative to which its clause is evaluated.¹⁹ The idea would then be that (certain) overt phrases, such as the prepositional phrase ‘in the family’ in (26a) as well as relatum DPs in functional descriptions such as (25), are an alternative way of specifying the value of this situation variable, making the overt article form unnecessary. Interestingly, there also seems to be some variation in the presence of the overt form corresponding to the difference between situational uniqueness through common knowledge vs. anaphoricity (27); however, much more work is needed to flesh out the full empirical picture here.

(27) Haitian Creole (Valdman 1977: 116)

- a. *Kote manje mwen?*
where meal my (interpreted relative to topic situation?)
- b. *Kote manje mwen an?*
where meal my DEF (based on previous mention)
‘Where is my meal?’

Theoretically, there are additional further implications of this type of approach as well. For example, global uniques would have to be assumed to require a situation pronoun (with a value distinct from the topic situation). Potentially interesting predictions arise with regards to intensional contexts, where situation pronouns fill the additional role of determining the intensional status of a given noun phrase (e.g. in terms of the *de re/de dicto* contrast). In this regard, the fact that *la* can occur on entire clauses as well would also be of further interest. And as already mentioned, the relationship between what happened to French-based *la* over time in Haitian and Mauritian Creole seems like a rich and important issue to explore. From the perspective just sketched, we might be dealing with a situation where the two take rather different paths to superficially similar but underlyingly distinct systems, roughly corresponding to the difference between representing anaphoric individual variables (as part of the strong article meaning) and representing variables for situations in the form of situation pronouns.

In sum, the case of Haitian Creole, which likely is mirrored in other languages as well, goes beyond what might be characterized as mere pragmatic variations in how the same meanings are put to use in the system of a given language, as reflected, e.g. in the lack of a bridging contrast in languages like Amern. A striking

¹⁹Note that the analysis of English demonstratives by Wolter (2006) develops some strikingly similar ideas for a different set of empirical facts.

observation, from the present perspective, is that even global uniques come with the overt form. The main question moving forward then will be to what extent the pattern represented here by Haitian Creole might reflect a fundamentally different type of contrast, or whether there are other languages that could be seen as further in-between cases, with a mix of the properties of the languages discussed in previous sections and cases like Haitian Creole. If the latter were the case, this might suggest that we are dealing with a more gradient spectrum after all, which would require some fairly substantial reconsiderations for an approach based on the formal article contrast as laid out above. I briefly review and comment on such a more gradient account in the following section.

5.2 Semantic vs. pragmatic uniqueness

A prominent alternative analysis goes back to Löbner (1985), with more recent developments in Löbner (2011) and, of particular relevance for our purposes, a fairly extensive typological discussion in Ortmann (2014). The core idea rests on a distinction between semantic and pragmatic uniqueness, which crucially rides on whether context has any role in establishing uniqueness. More specifically, semantic uniqueness holds if a definite description refers unambiguously based on the meaning of the noun alone, in a context-independent manner. In contrast, in cases of pragmatic uniqueness, reference is unambiguous only under consideration of contextual information, which can be linguistic or extra-linguistic. Crucially, this distinction is seen relative to a gradient uniqueness scale, which allows different languages to choose different cut-off points for using one form as opposed to another. Ortmann (2014) succinctly states the role of these notions for article contrasts (or “splits”):

[...] the distinction between semantic and pragmatic uniqueness is the basis of all conceptually governed article splits, in that a shift towards an IC [Individual Concept] or FC [Functional Concept] is overtly signaled.

(Ortmann 2014: 296)

The approach crucially rests on the assumption that nouns differ lexically from one another with regards to their semantic types. Table 3 provides an overview of the key dimensions of variation, namely a) whether their meanings are at their core referential (ending in type *e*) or predicative (functions from a given number of individuals to truth values).

However, the type of nouns can be adjusted through (fairly standard) type-shifting operations. Definite noun phrases are generally analyzed as functional concepts, in that they are assumed to refer unambiguously. However, that status is attained in different ways, in that some nouns require a type-shifter, and others do not. The difference between two distinct definite articles is then captured

1 Weak vs. strong definite articles: Meaning and form across languages

Table 3: Semantic vs. pragmatic uniqueness (adapted from Ortman 2014)

	Monadic	Polyadic
Non-unique (pragmatic)	Sortal nouns <i>dog, stone</i> $\langle e, t \rangle$	Relational nouns <i>sister, finger</i> $\langle e, \langle e, t \rangle \rangle$
Unique (semantic)	Individual nouns <i>sun, prime minister</i> $\langle e \rangle$	Functional nouns <i>father, head</i> $\langle e, e \rangle$

in terms of the signal they convey about how uniqueness was achieved. For example, the idea for Standard German would be that the strong article indicates pragmatic uniqueness, whereas the weak article indicates semantic uniqueness.

This idea is made more flexible by the notion that different types of noun phrases relate to the context in different ways. Based on this, the approach assumes a scale of uniqueness, “defined according to the degree of invariance of reference of nominal expressions” (Ortman 2014):

- (28) Scale of uniqueness (Ortman 2014: 314; adapted from Löbner 2011)
 deictic sortal noun < anaphoric sortal noun < SN with establishing relative clause < relational Definite Associative Anaphora* < part-whole Definite Associative Anaphora, non-lexical functional nouns, < lexical individual nouns/functional nouns < proper names < personal pronouns

Essentially, a language with a contrast between definite articles could then draw the line anywhere on this scale, marking expressions to one side with a weak article and those to the other side with the strong article. Intuitively, the idea is that different nouns require different amounts of lifting to end up with the right semantic type for a definite description, and the articles serve as indicators of whether a certain amount of lifting had to occur. The approach naturally affords a substantially more fine-grained set of typological options than any simple binary contrast.

While not all relevant aspects of this proposal can be discussed here, let us briefly assess both challenges and strengths of this general approach.

Starting with the former, there is a question at the level of the general architecture of the syntax-semantics interface with regards to the mapping from syntactic categories to semantic types. While it is clear that we have to allow for some flexibility, e.g. with regards to the number of arguments a given predi-

cate involves, sub-dividing the space of lexical entries for nouns into predicates and entities gives rise to additional complications. These are by no means insurmountable, but their repercussions have to be assessed carefully. On the flipside of the coin, determining the availability of the type-shifters that are standardly invoked for dealing with these complications has to be carefully constrained. Another aspect that requires further spelling out is the nature of the measure on the uniqueness scale, especially as new potential contrasts are considered based on new data from additional languages. On the semantic side, the question arises of how cases where there is a clear overall meaning contrast based on which article is used are captured in the formal derivation if the articles themselves do not contribute any meaning. Finally, the specification of the key notions of uniqueness tries to characterize unambiguous reference relative to the denotation of the noun (since it is based on lexical properties), rather than the full noun phrase. But this does not translate straightforwardly to cases of more complex noun phrases, where traditional uniqueness-based analyses crucially rely on the compositional combination of the determiner with its complement noun phrase as a whole (e.g. including modifying adjectives). Relatedly, it is not obvious how the broader integration of this approach into a formal semantic system that interacts with the grammar should proceed, specifically with regards to the various mechanisms for co-variation under quantifying expressions briefly discussed above.

There are empirical problems for this type of approach as well. In particular, sortal nouns of various kinds can be turned functional through appropriate contexts – as illustrated by the following variation on (7b) (where a strong article was required):

(29) German

Context: Hans, who works at a ministry, and his wife are talking about what has been going on at work.

a. What happened to the proposal you drafted?

b. *Der Vorschlag wurde in der Kabinettsitzung gestern vom_s*
the proposal was in the cabinet meeting yesterday by_the_{weak}
Minister vorgestellt, aber 7 SPD-Minister haben dagegen gestimmt.
minister introduced but 7 SPD-ministers have against voted

‘The proposal was introduced by the minister in yesterday’s cabinet meeting, but 7 SPD-ministers voted against it.’

Crucially, nothing about the noun in such cases ensures uniqueness directly, and to the extent that uniqueness does hold, that only is so based on a substantial

1 Weak vs. strong definite articles: Meaning and form across languages

amount of contextual information – in essence, the entire definite noun phrase is interpreted relative to the speaker’s work place here. But surely such a contextual modulation should not lead us to consider different lexical entries for the word ‘minister’.²⁰

Let us now turn to some of the strengths of this proposal. First, as already noted above, it allows for a substantial range of variation between languages along a single dimension, and Ortmann (2014) applies the resulting prediction in interesting ways, both synchronically and diachronically. But even as that success should be registered, it is worth noting that the formal proposal on its own predicts that languages should be able to choose a cut-off point anywhere on the scale. In light of the variation present in existing data, it seems that even though some flexibility is needed, the full range of options goes beyond what is required (of course this could change with additional data being brought under consideration).

In relation to these concerns, it is also worth revisiting some aspects of Haitian Creole in light of the analysis in terms of semantic vs. pragmatic uniqueness. The uniqueness scale has global uniques on par with functional nouns with explicit arguments. But Haitian Creole crucially draws a line between these two, and any plausible additional split of the uniqueness scale would predict an opposite ordering from what is empirically attested in this regard. Furthermore, the intriguing interaction of *la* with the domain of *only* would not seem to be something that can be explained in any straightforward way from this perspective.

In sum, accounts based on the distinction between semantic and pragmatic uniqueness do have some desirable empirical predictions going for them, but they also face some challenges, both conceptually and theoretically. In light of this, it should be clear that accounting for the full range of article variation across languages requires substantially more work, regardless of the theoretical approach one starts out with. But the empirical picture overall is not incompatible with a view where the core weak vs. strong contrast is mirrored in properties of article contrasts across many languages, but various other, potentially independent, factors can affect just what form is thought to be ideally suited for the purposes at hand.

²⁰Note also that this is clearly a different contrast than that in the sketch of Haitian Creole above, where resource situation would require a strong article.

6 Conclusion

In this chapter, I have reviewed the key tenets of the contrast between weak and strong article definites presented in Schwarz (2009), and considered a range of data across various languages in light of it. There seems to be a substantial number of languages from entirely unrelated language families that use different forms for different types of definite noun phrases in a way that seems to reflect the weak vs. strong article contrast found in Germanic. While there are some minor variations in the pragmatics of which forms get used when both are available, the nature of the semantic contrast in a large set of languages seems to be fairly uniform and consistent with an analysis in terms of situational uniqueness and anaphoricity. In addition, the formal realization of the contrasts was considered, and there is at least preliminary evidence from the languages discussed that there is real variation in the interpretation of bare noun phrases, in a way that suggests that distinct null D-heads may be at play in at least some of them.

Additional languages enriched the picture further, as they exhibit contrasts that clearly seem to go beyond the weak vs. strong contrast. There are two possible approaches to tackling this. First, one can see these languages in terms of orthogonal factors, providing insights into potentially related, but ultimately separate dimensions of variation. Alternatively, one can see them in terms of a more gradient perspective on how different types of definites are signaled within a grammar, as on the approach based on semantic vs. pragmatic uniqueness. Both types of approaches require extensions and elaborations, so more work is needed both empirically and theoretically to achieve a more conclusive assessment of the semantic typology of definiteness across languages. However, the sharpening of key descriptive notions and crucial contrasts goes a long way towards having more precise tools that can help to get a more uniform and broad cross-linguistic perspective on the nature and extent of variation.

References

- Ahn, Dorothy. 2016. Three kinds of definites. (Cambridge: Harvard University. Unpublished manuscript).
- Ahn, Dorothy. 2017. Three kinds of definites: Weak, strong, and deictic. (Poster presented at Sinn und Bedeutung 21, Edinburgh, 4–6 September 2016).
- Arkoh, Ruby & Lisa Matthewson. 2013. A familiar definite article in Akan. *Lingua* 123. 1–30.

1 *Weak vs. strong definite articles: Meaning and form across languages*

- Birner, Betty & Gregory Ward. 1994. Uniqueness, familiarity, and the definite article in English. In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, vol. 1, 93–102. Berkeley: Berkeley Linguistics Society.
- Bombi-Ferrer, Carla. 2017. *Definiteness in Akan*. Potsdam: University of Potsdam. (MA thesis).
- Brugger, Gerhard & Martin Prinzhorn. 1996. Some properties of German determiners. (Vienna: University of Vienna. Unpublished manuscript).
- Cabredo Hofherr, Patricia. 2013. Reduced definite articles with restrictive relative clauses. In Patricia Cabredo Hofherr & Anne Zribi-Hertz (eds.), *Crosslinguistic studies on noun phrase structure and reference* (Syntax and Semantics 39). Leiden: Brill.
- Carlson, Greg N., Rachel Shirley Sussman, Natalie M. Klein & Michael K. Tanenhaus. 2006. Weak definite noun phrases. *Proceedings of NELS* 36(1). 179–196.
- Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6(4). 339–405.
- Cho, Jacee. 2016. The acquisition of different types of definite noun phrases in L2-English. *International Journal of Bilingualism* 21(3). 367–382.
- Christophersen, Paul. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Einar Munksgaard.
- Cieschinger, Maria. 2006. *Constraints on the contraction of preposition and definite article in German*. Osnabrück: University of Osnabrück. (BA Thesis).
- Clark, Herbert H. 1975. Bridging. In *TINLAP '75 Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing*, 169–174. Stroudsburg: Association for Computing Machinery.
- Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4). 393–450.
- Dayal, Veneeta. 2016. Semantic universals: The case of definites and generics. (Talk presented at the Panel on Universals at the 40th Annual Penn Linguistics Conference, Philadelphia, 18–20 March 2016).
- Ebert, Karen. 1971a. *Referenz, Sprechsituation und die bestimmten Artikel in einem Nordfriesischen Dialekt (Fering)*. Kiel: Christian-Albrechts-Universität zu Kiel. (Doctoral dissertation).
- Ebert, Karen. 1971b. Zwei Formen des bestimmten Artikels. In Dieter Wunderlich (ed.), *Probleme und Fortschritte der Transformationsgrammatik*, 159–174. München: Hueber.
- Elbourne, Paul D. 2013. *Definite descriptions*. Oxford: Oxford University Press.
- Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik* 100. 25–50.

- Grove, Julian & Emily Hanink. 2016. Article selection and anaphora in the German relative clause. *Proceedings of SALT* 26. 417–432.
- Grubic, Mira. 2016. Interaction of definite & indefinite determiners in Ngamo (W-Chadic). (Talk presented at the Workshop on the Semantic Contribution of Det and Num: (In)definiteness, genericity and referentiality, Barcelona, Universitat Autònoma de Barcelona, 2016).
- Haberland, Hartmut. 1985. Zum Problem der Verschmelzung von Präposition und bestimmtem Artikel im Deutschen. *Osnabrücker Beiträge zur Sprachtheorie* 30. 82–106.
- Hanink, Emily. 2017. The German definite article and the ‘sameness’ of indices. *University of Pennsylvania Working Papers in Linguistics* 23(1). 63–72.
- Hartmann, Dietrich. 1967. *Studien zum bestimmten Artikel in ‘Morant und Galie’ und anderen rheinischen Denkmälern des Mittelalters*. Giessen: Wilhelm Schmitz Verlag.
- Hartmann, Dietrich. 1978. Verschmelzungen als Varianten des bestimmten Artikels? In Dietrich Hartmann, H.-J. Linke & O. Ludwig (eds.), *Sprache in Gegenwart und Geschichte. Festschrift für Heinrich Matthias Heinrichs*. 68–81. Köln: Böhlau.
- Hartmann, Dietrich. 1980. Über Verschmelzungen von Präposition und bestimmtem Artikel. Untersuchungen zu ihrer Form und Funktion in gesprochenen und geschriebenen Varietäten des heutigen Deutsch. *Zeitschrift für Dialektologie und Linguistik* 47. 160–183.
- Hartmann, Dietrich. 1982. Deixis and anaphora in German dialects: The semantics and pragmatics of two definite articles in dialectal varieties. In Jürgen Weissenborn & Wolfgang Klein (eds.), *Here and there: Cross-linguistic studies on deixis and demonstration* (Pragmatics and Beyond III: 2-3), 187–207. Amsterdam: John Benjamins.
- Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croon Helm.
- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. Amherst: University of Massachusetts. (Doctoral dissertation).
- Heinrichs, Heinrich Matthias. 1954. *Studien zum bestimmten Artikel in den germanischen Sprachen*. Giessen: Wilhelm Schmitz Verlag.
- Herburger, Elena. 1997. Focus and weak noun phrases. *Natural Language Semantics* 5(1). 53–78.
- Himmelman, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Tübingen: Niemeyer.

- Ingason, Anton. 2016. *Realizing morphemes in the Icelandic noun phrase*. Philadelphia: University of Pennsylvania. (Doctoral dissertation).
- Irani, Ava. 2016. Two types of definites in American Sign Language. (Talk presented at the workshop of Definiteness Across Languages, Mexico City, 22–24 June 2016).
- Irani, Ava. 2019. On (in)definite expressions in American Sign Language. In Ana Aguilar-Guevara, Julia Pozas & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages*. Berlin: Language Sciences Press.
- Irani, Ava & Florian Schwarz. 2016. Two types of definites in American Sign Language. (Talk presented at the LSA 2016 Annual Meeting, Washington D.C., 7–10 January 2016).
- Jenks, Peter. 2015. Two kinds of definites in numeral classifier languages. *Proceedings of SALT* 25. 103–124.
- Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal methods in the study of language*, vol. 1, 277–322. Amsterdam: Mathematical Centre Tracts.
- Koulidobrova, Elena V. 2012. *When the quiet surfaces: ‘Transfer’ of argument omission in the speech of ASL-English bilinguals*. Storrs: University of Connecticut. (Doctoral dissertation).
- Leu, Thomas. 2008. *The internal syntax of determiners*. New York: New York University. (Doctoral dissertation).
- Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4(4). 279–326.
- Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28(3). 279–333.
- Lyons, Christopher. 1999. *Definiteness* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Milsark, Gary L. 1977. Towards the explanation of certain peculiarities of existential sentences in English. *Linguistic Analysis* 3. 1–29.
- Neale, Stephen. 1990. *Descriptions*. Cambridge: MIT Press.
- Ortmann, Albert. 2014. Definite article asymmetries and concept types: Semantic and pragmatic uniqueness. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald & Wiebke Petersen (eds.), *Frames and concept types: Applications in linguistics and philosophy* (Studies in Linguistics and Philosophy 94), 293–321. Berlin: Springer.
- Partee, Barbara H. 1986. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh & Martin Stokhof (eds.), *Studies in discourse representation theory and the theory of generalized quantifiers*, 115–143. Dordrecht: Foris.

- Percus, Orin. 2000. Constraints on some other variables in syntax. *Natural Language Semantics* 8(3). 173–229.
- Poesio, Massimo. 1994. Weak definites. *Proceedings of SALT* 4. 282–299.
- Poesio, Massimo & Renata Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics* 24(2). 183–216.
- Prince, Ellen F. 1981. Toward a taxonomy of given/new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Puig Waldmüller, Estela Sophie. 2008. *Contracted preposition-determiner forms in German: Semantics and pragmatics*. Barcelona: Universitat Pompeu Fabra (Doctoral dissertation).
- Reimer, Marga & Anne Bezuidenhout (eds.). 2004. *Descriptions and beyond*. Oxford: Oxford University Press.
- Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26(3). 287–350.
- Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493.
- Scheutz, Hannes. 1988. Determinantien und Definitheitsarten im Bairischen und Standarddeutschen. In *Festschrift für Ingo Reiffenstein zum 60. Geburtstag*, 231–258. Göppingen: Kümmerle.
- Schlenker, Philippe. 2017. Sign language and the foundations of anaphora. *Annual Review of Linguistics* 3. 149–177.
- Schmitt, Viola. 2006. *Hessische Relativsätze*. Vienna: University of Vienna. (MA thesis).
- Schuster, Mauriz & Hans Schikola. 1984. *Sprachlehre der Wiener Mundart*. Wien: Österreichischer Bundesverlag.
- Schwager, Magdalena. 2007. (Non-)functional concepts: Definite articles in Bavarian. (Talk presented at the 8th Szklarska Poreba Workshop, Szrenica, 23–27 February 2007).
- Schwarz, Florian. 2009. *Two types of definites in natural language*. Amherst: University of Massachusetts. (Doctoral dissertation).
- Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559.
- Šereikaitė, Milena. 2016. Definiteness in Lithuanian. (Philadelphia: University of Pennsylvania. Manuscript).
- Šereikaitė, Milena. 2019. Strong vs. weak definites: Evidence from Lithuanian adjectives. In Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages*. Berlin: Language Science Press.

- Šimík, Radek. 2015. On pragmatic uses of demonstratives: The case of pragmatic discourse anaphora in Czech. (Potsdam: University of Potsdam. Manuscript).
- Simonenko, Alexandra. 2014. *Grammatical ingredients of definiteness*. Montreal: McGill University. (Doctoral dissertation).
- Simpson, Andrew. 2017. Bare classifier/noun alternations in the Jinyun (Wu) variety of Chinese and the encoding of definiteness. *Linguistics* 55(2). 305–332.
- Simpson, Andrew & Priyanka Biswas. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis* 40(3–4). 167–198.
- Stanley, Jason & Zoltán Gendler Szabó. 2000. On quantifier domain restriction. *Mind and Language* 15(2). 219–261.
- Strawson, Peter F. 1950. On referring. *Mind* 59(235). 320–344.
- Valdman, Albert. 1977. *Pidgin and creole linguistics*. Bloomington: Indiana University Press.
- von Fintel, Kai. 1994. *Restrictions on quantifier domains*. Amherst: University of Massachusetts. (Doctoral dissertation).
- Wespel, Johannes. 2008. *Descriptions and their domains: The patterns of definiteness marking in French-related creoles*. Stuttgart: University of Stuttgart. (Doctoral dissertation).
- Westerståhl, Dag. 1984. Determiners and context sets. In Johan van Benthem & Alice G. B. ter Meulen (eds.), *Generalized quantifiers in natural language*, 45–71. Dordrecht: Foris.
- Wiltschko, Martina. 2013. Descriptive relative clauses in Austro-Bavarian German. *Canadian Journal of Linguistics* 58(2). 157–189.
- Wolter, Lynsey Kay. 2006. *That's that: The semantics and pragmatics of demonstrative noun phrases*. Santa Cruz: University of California. (Doctoral dissertation).

