Article Use Across Languages: an OT Typology[*]

Henriëtte de Swart
Dept. of French/UiL-OTS
Utrecht University

Henriette.deswart
@let.uu.nl

Joost Zwarts
Dept. of Ling./UiL-OTS
Utrecht University

Joost.Zwarts
@let.uu.nl

Abstract

Many languages use definite and indefinite articles to signal the referential status of individuals in conversational space. But articles are not necessary to set up or maintain discourse reference, as witnessed by the fact that languages can also get by with bare nominals. Similarly, we find that many, but not all languages use a morphological singular/plural contrast to distinguish reference to atoms from reference to sums. We assume that an awareness of discourse reference and atomic/sum reference is part of universal human cognition. The balance between economy considerations (favoring bare nominals) and the desire to convey (discourse) referential distinctions determines whether these cognitive operations are reflected in the form nominals take in the language. A range of possible OT grammars is discussed, and exemplified. The cross-linguistic semantics of bare nominals is related to these grammars. Bare nominals have the interpretations that are not blocked by other, more specific forms.

1   Introduction

In languages like English, we use articles to set up referents in a conversational space, and to refer back to them (Kamp 1981, Heim 1982). Consider the examples in (1).

(1)   a.   A student$_i$ came to my office. She$_i$ had a question about the exam.

　　b.　　A child$_i$ was playing in the park. The funny little creature$_i$ wore a green hat and purple socks.

The indefinite noun phrase *a student* (1a) sets up a referent in the universe of the discourse. This discourse referent has the properties of being a student and of having come to my office. Once a referent with such properties has been introduced, we can use a pronoun to refer back to it. Thus *she* in the subsequent sentence of (1a) is an anaphoric pronoun that refers back to the student who came into my office. Instead of a pronoun, we can also use a definite description to refer to a familiar discourse referent (1b). Definites convey uniqueness (singulars, examples in 1) or maximality (plurals, examples in 2). In anaphoric contexts such as (1), uniqueness or maximality is relativized to the discourse. In other contexts, world knowledge plays a role, as in (2).

(2)　　　I love you more than the sun and the stars.

These informal descriptions of (1) and (2) imply that definite and indefinite articles play very specific roles in conversation, which are best evaluated at the discourse level. Rijkhoff (2002: 185) states that definite articles are localizing elements that express 'weak deixis' in the sense of Anderson and Keenan (1985: 261-2). Farkas (2002) develops a notion of dynamic uniqueness in Discourse Representation theory (DRT), which is labeled as 'determined reference'. Determined reference generalizes over uniqueness/maximality and anaphoricity, and requires that the value chosen for a discourse referent introduced by a definite NP at a certain point of update is fixed. All possible update functions extending the input context assign to the discourse referent the same individual in the model. Nominals with indefinite articles lack these features, and simply introduce a discourse referent (Kamp 1981, Kamp and Reyle 1993).

In languages like English, articles (and determiners more generally) are the prime expressions introducing discourse referents (cf. Kamp and Reyle 1993, Higginbotham 1985, Kamp and van Eijck 1997). The contrast between bare singulars and bare plurals in (3) suggests that plural morphology also plays a dynamic role.

(3)　　　Mary ate *apple/an apple/the apple/apples/some apples/the apples.

Farkas and de Swart (2003) interpret plural morphology in terms of a predication of plurality over a discourse referent. In order to support such a predication, there must be a discourse referent available. Plural morphology thus presupposes a discourse referent. This presupposition can be bound by a determiner (*the, some*, etc.) or accommodated, which leads to the bare plural in (3).

So far, our examples have been chosen from English. Other languages that have articles or singular/plural morphology use these linguistic means to convey very similar meaning distinctions. This leads us to posit the following correspondence rules between forms and meanings:

- ♦ **Dr:** A determiner (form) corresponds with the presence of a discourse referent (meaning).

- ♦ **Def:** A definite article (form) corresponds with a discourse referent with determined reference (meaning).

- ♦ **Pl:** Plural morphology on the noun (form) corresponds with a predication of plurality on a presupposed discourse referent (meaning).

The correspondence between forms and meanings is quite stable across languages, but clearly, not all languages have definite/indefinite articles, or establish a singular/plural distinction in the morphology of the noun. Many languages do not mark the contrast between sum/atomic reference on the noun, and/or use *bare nominals*, i.e. noun phrases that lack an article or determiner in environments in which English would use a definite or indefinite noun phrase. Depending on the language, such a bare nominal may have a definite or an indefinite, a singular or a plural meaning. (4) illustrates this for Mandarin Chinese (the example is from Krifka 1995).

(4)   Wò   kànjiàn   xióng   le
           I   see   bear   ASP
           'I see a bear/some bears/the bear(s)'

We assume that setting up referents in discourse space, and referring to them involve general cognitive operations, which are related to the way human beings organize the conversational space around the individuals that we talk about. Furthermore, we take everyone to possess the same cognitive abilities, independently of their mother tongue. In the face of such assumptions about general cognitive abilities, linguists wonder why certain languages (such as English) necessarily encode these cognitive operations in the functional structure of nominals, whereas other languages (such as Mandarin Chinese) do not. Proposals addressing this issue appeal to covert functional structure (Borer 2004) or parametric variation (Chierchia 1998a). These proposals have generated many new insights, but also face empirical and conceptual problems. For lack of space, we will not discuss these proposals here, but focus on an alternative line of explanation in terms of Optimality Theory (OT). We posit the correspondence rules **Dr, Def** and **Pl** as universal constraints on the mapping from meanings to forms. This implies that languages that have definite articles or a singular/plural contrast assign roughly the same meaning to these forms. However, languages do not need to use these forms to convey the relevant meanings. Language variation in the expression of plurality, discourse reference, and determined reference arises from the interaction of the mapping from meaning to form with a general economy constraint blocking functional structure in the nominal domain. Such an economy constraint favors bare nominals and nominals not marked for number over nominals with an elaborate functional structure. This economy constraint is also a rule of universal grammar, but it obviously conflicts with

the correspondence rules. The conflict is resolved by taking the correspondence rules from form to meaning to be soft, violable constraints, that are ranked with respect to each other, in an Optimality Theoretic way. The ranking is language-specific, and a range of possible grammars arises from this interaction. This idea is developed in Section 2. We know that bare nominals do not have the same meaning from one language to the next. In the absence of form carrying meaning, the correspondence rules do not have anything to say about the semantics of bare nominals. This leads to the idea that bare nominals can only have those meanings that are not blocked by more specific nominals in the language. Section 3 develops a cross-linguistic semantics of bare nominals that relies on the OT typology developed in Section 2.

2   Markedness and faithfulness in the nominal domain

As far as the syntactic structure of nominals is concerned, we assume that full DPs have the layered structure $[_{DP} [_{NumP} [_{NP}]]]$. Articles, demonstratives, numerals and quantifiers reside in D. We try to avoid null elements in the functional projections Num and D in teh absence of evidence that they need to be projected. Whether nominals project at the functional levels of NumP and DP depends on the presence of number morphology on the noun and on the presence of articles or other determiners. All nominals project at least at the lexical level of NP. Thus nominals can have the structure of an NP (bare singulars), a NumP (bare plurals) or a DP (nominals with an article or some other determiner). Bare nominals constitute the unmarked form, because they have the simplest possible nominal structure. The OT grammar values this insight by means of a core markedness constraint, viz. *FunctN, which avoids all functional structure in the nominal domain:

◆   *FunctN: Avoid functional structure in the nominal domain.

*FunctN is motivated by economy: less functional structure is better than more. *FunctN is a gradable constraint, and each functional projection in the noun phrase presents a violation of the markedness constraint. In this section, we will present six classes of languages (labeled i-vi), based on six grammars in which the economy constraint *FunctN interacts with faithfulness constraints driving the expression of atomic/sum reference and discourse referential information. Reranking of constraints is driving the typology in an optimality-theoretic fashion.

If the constraint *FunctN is ranked higher than faithfulness constraints involving the expression of meanings that are characteristically expressed in the functional layer above NP (call them $FNom_{a...z}$), we obtain the ranking *FunctN >> {$FNom_a$, $FNom_b$, ... $FNom_z$}. Under this ranking, we have no singular/plural distinctions, no articles, no indefinite determiners (like *some, several*), no numerals (like *four, at least three*), and no D-quantifiers (like *every, most*). The mutual ranking of the constraints $FNom_a$, $FNom_b$, ... $FNom_z$ is irrelevant if all functional structure is blocked by the highly ranked

markedness constraint *FunctN. Arguably, this ranking would give us the perfectly unmarked nominal system. From the description of a wide range of languages as provided by the literature, we learn that most if not all natural languages have at least some level of functional structure in the nominal domain. Even if they don't have morphological number and do not use articles, they may have case marking, numerals, classifiers, demonstratives, quantifiers, or some other kind of functional structure. Example (5) (from Chierchia 1998b) shows this for Mandarin Chinese.

(5)   liǎng   zhāng   zhuōzi
        two   CL   table
        'two (pieces of) table'

For reasons of space, we cannot address the full range of functional structure in this paper. The contrast between articles and other types of determiners suggests that articles have a special status among determiners. It is not always easy to determine which expressions living in D qualify as articles, and which do not. In the context of this paper, we take articles to be determiners that introduce a discourse referent (cf. the correspondence rule **Dr**), and possibly convey determined reference (cf. the correspondence rule **Def**), but do not contribute other quantificational/qualitative information. In order to reflect the special status of articles in the OT system, we formulate the constraint *Art that can be viewed as a special instance of the general markedness constraint *FunctN:

♦   *Art: Avoid article.

If *Art and *FunctN are ranked above faithfulness constraints governing article use and a reflection of the singular/plural distinction (collectively labeled as {$FNom_1$, $FNom_2$, ... $FNom_n$}, and to be spelled out in the remainder of this section), we don't see a formal reflection of plurality or information concerning discourse referential status conveyed by articles. Bare forms are used for definite, indefinite and kind reference, and do not display a singular/plural distinction, as illustrated by the Mandarin Chinese example (4) above. We may posit other faithfulness constraints not discussed in this article (collectively labelled as $FNom_x$) to reside below *Art, but above *FunctN, and which give rise to functional structure including demonstratives, classifiers, quantifiers, etc (cf. Example 5). Such languages as Japanese, Thai and Mandarin Chinese are then class (i) languages with the ranking *Art >> $FNom_x$ >> *FunctN >> {$FNom_1$, $FNom_2$, ... $FNom_n$}. Given that we do not address the nature of the constraints $FNom_x$ in this paper, we will leave them out of the rankings that build up the typology in this section.

Mandarin Chinese is a language without a formal reflection of the singular/plural distinction. But in many languages, number is marked in the morphology on the noun. As a default, plural will be marked first, if there is a number distinction at all (Greenberg 1963, Corbett 2000). In line with these observations, we posit a faithfulness constraint FPL, which picks up on the correspondence rule **Pl** formulated in Section 1.

◆   FPL: Plural predication on a (presupposed) discourse referent corresponds
                 with an expression in Num.

Languages that rank FPL higher than *FunctN project at least a NumP for plural nouns. However, the expression of number in the morphology on the noun is not necessarily connected to article use. Languages that adopt the ranking FPL >> {*Art, *FunctN} >> {Fnom$_1$, Fnom$_2$, ... Fnom$_n$}, establish a formal distinction between singular and plural nouns, but definite and indefinite meanings as well as kind reference is expressed by bare nominals, i.e. nominals without an article (singulars and/or plurals). Relevant class (ii) languages are Hindi, Russian, Georgian. (6) provides an example from Georgian (from Harris (1981: 21-22, quoted by Rijkhoff 2002). The Russian example (7) is from Chierchia (1998a); Dayal (2004) discuss similar examples from Hindi.

(6)   burtebi   goravs
         balls:PL:NOM   roll:3SG
         'Balls/the balls are rolling.'

(7)   V   komnate   byli   malcik   I   devocka.
         In room   were   boy   and   girl
         'In the room were a boy and a girl.'

Although the high ranking of *Art and *FunctN in class (i) and class (ii) languages is economical from the production perspective (it reduces speaker effort), it is not attractive from the comprehension perspective (it induces massive ambiguities for the hearer to resolve). As has been shown in the literature, bare nominals in Manarin Chinese, Hindi and Georgian lend themselves to definite, existential, and generic interpretations. Case marking, word order, lexical/prosodic information and other contextual clues may be exploited by the hearer to construct the optimal discourse interpretation of the nominal. But the interpretation process would be facilitated if some of these meaning distinctions would be reflected in the form of the nominal. This is the intuition underlying the faithfulness constraints FDEF and FDR, which mirror the correspondence rules **Def** and **Dr** introduced in Section 1 above.

FDEF requires the expression of determined reference. The prototypical way to realize determined reference is by means of a definite determiner, and for simplicity we will build this into the constraint.

◆   FDEF: Determined reference of a discourse reference corresponds with an
                  expression of definiteness in D.

With the ranking FDEF >> *Art, we obtain a system that exemplifies an alternation between definites and bare nominals. Such class (iii) languages are Hebrew and

Bulgarian. The following Hebrew examples illustrate (8 is from de Swart, Winter and Zwarts 2007; 9 is from Doron 2004).

(8)   dan   ra'a   namer
        Dan   saw   tiger
        'Dan saw a tiger.'

(9)   namer/ ha-namer   Hu   xaya   Torefe
        tiger   the-tiger Is   animal   carnivorous
        'The tiger is a carnivorous animal.'

The bare singular gets an existential (8) or a generic interpretation (9). The definite singular gets a regular definite or a generic interpretation (9). Class (iii) languages have a stable system that reflects the asymmetry between definite and indefinite (or rather: non-definite) nominals defended in Farkas (2002, 2006) and Farkas and de Swart (2007, 2008), who maintain that the definite article marks determined reference, whereas non-definites are simply unmarked. If we posit FDEF as a relevant faithfulness constraint, we predict that languages typically oppose bare and definite nominals, but not bare and indefinite ones. This is in line with the typological literature (Greenberg et al. 1978). The semantic literature has emphasized that bare nominals (singular and plural), in languages in which they occur, always have an existential meaning (possibly besides definite and generic readings). The availability of an existential interpretation for bare nominals is common to class (i) to (iii) languages.

The introduction of faithfulness constraints concerning the referential status of the discourse referent should be viewed in the broader perspective of how discourse referents are introduced. Farkas and de Swart (2003) develop a more fine-grained version of Kamp and Reyle's (1993) Discourse Representational Theory, and introduce a distinction between discourse referents and so-called thematic arguments. Thematic arguments are the argument slots associated with nouns and verbs as they come out of the lexicon. Thematic arguments are part of DRS conditions, but are not members of the universe of discourse of the DRS, so they do not have dynamic force. At the compositional level of building predicate-argument structure, these thematic arguments must be instantiated by discourse referents. Farkas and de Swart (2003) develop modes of composition that place the burden of contributing discourse referents on the nominal. More precisely, the determiner introduces a discourse referent which instantiates the thematic argument of the noun in building the DP. When a DP in regular argument position combines with the verb, its discourse referent instantiates the thematic argument of the verb. A dynamic semantics of plural morphology allows bare nominals to introduce a discourse referent via accommodation of the presupposition introduced by the plural morphology (cf. Section 1). Not all languages allow this accommodation process, so we establish a distinction between "strong" plural morphology permitting to bare plurals in regular argument position (as in Germanic languages), and "weak" plural

morphology incapable of licensing a discourse referent in regular argument position (as in Salish, French, see below). In accordance with this view, we posit a constraint that requires discourse referents to be introduced by a strong functional layer above the NP. We label this constraint FDR.

◆   FDR: The presence of a discourse referent in the semantics corresponds with a strong functional layer above NP.

Languages that rank FDR below *FUNCTN don't require a functional layer above NP to parse discourse referents, so bare nominals are fully adequate in regular argument positions, as is commonly established for class (i) to (iii) languages. Class (iv) languages that adopt the ranking {FPL, FDR} >> {*Art, *FunctN} >> FDEF do not establish a definite/indefinite contrast, but use a determiner on all nominals in argument position. The Salish languages (a family of Amerindian languages spoken in Canada) exemplify this ranking. Matthewson (1998) provides examples from St'át'imcets, and points out that all argument nominals are introduced by an overt determiner (10).

(10)   tecwp-mín-lhkan   ti púkw-a   lhkúnsa
          buy.APPL-1SG.SUB   DET book-DET   today
          'I bought a/the book today.'

According to Matthewson, the St'át'imcets determiners do not encode either definiteness or specificity, but 'assertion of existence'. She models this notion in DRT (Kamp and Reyle 1993). For our purposes, the notion of 'assertion of existence' can be identified with the introduction of a discourse referent. In St'át'imcets, morphological number is incapable of licensing a discourse referent, so we take this language to have a "weak" plural. Accordingly, the presence of an overt determiner is required, even in the plural (11) (cf. below for a treatment of French along similar lines).

(11)   léxlex   i   smelhmúlhats-a
          intelligent   DET.PL   woman(pl)-DET
          'Women/the women are intelligent.'

The ranking {FPL, FDR} >> {*Art, *FunctN} >> FDEF postulated for St'át'imcets derives full DPs for inputs containing a discourse referent.

Languages that rank FPL, FDEF and FDR above *Art and *FunctN exemplify a full contrast between definite and indefinite forms. In line with Farkas and de Swart (2003), we assume that plural morphology on the noun may also be capable of licensing a discourse referent (see above). Assuming that singular nouns do not have a Num projection, but plural nouns do, we end up with the asymmetry between singular and plural nouns illustrated in (3) for English. Bare plurals occur in regular argument position, but bare singulars are blocked. If either Num or D is sufficient to introduce a discourse referent, class (v) languages end up with a system of definite and indefinite

articles in the singular, and an opposition between definites and bare nominals in the plural. This is the system characterizing English (but also Dutch, German, Norwegian, Swedish, Spanish, Italian). We illustrate this for Dutch in (12).

(12)   Sanne   kocht   *appel/   een appel/   de appel/   appels/   de appels
         Sanne   bought *apple/   an apple/   the appel/ apples/   the apples
         'Sanne bought *apple/an apple/the apple/apples/the apples.'

Tableaux 1 and 2 illustrate how the ranking {FPL, FDR, FDEF} >> {*Art, *FunctN} derives the possibility of bare plurals, but not of bare singulars in regular argument position (i.e. the variable *x* in the input stands for a discourse referent).

Tableau 1: blocking of bare singular in class (v) languages, illustrated with English (production).

Meaning $\exists x$ Book(x)	Form	FPL	FDR	FDEF	*Art	*FunctN
	[$_{NP}$ book]		*			
	[$_{NumP}$ sg [$_{NP}$ book]]		*			*
☞	[$_{DP}$ a [$_{NumP}$ sg [$_{NP}$ book]]]				*	**

Tableau 2: bare plurals in class (v) languages, illustrated with English (production).

Meaning $\exists x$ Book(x) & Pl(x)	Form	FPL	FDR	FDEF	*Art	*FunctN
	[$_{NP}$ book]	*	*			
☞	[$_{NumP}$ pl [$_{NP}$ book]]					*
	[$_{DP}$ indef [$_{NumP}$ pl [$_{NP}$ book]]]				*	**

In English-type languages, bare singulars are blocked, because of the ranking of FDR > {*Art, *FunctN}, as Tableau 1 illustrates. It does not help to project a NumP, because there is no singular morphology in English that is licenses a discourse referent. The only way to satisfy FDR with a singular, indefinite meaning in the input is to project an indefinite article *a* in D. As we see in Tableau 2, plural morphology in English is strong and satisfies FDR. The insertion of a plural indefinite article would constitute an unnecessary violation of *Art and *FunctN. The contrast between Tableaux 1 and 2 illustrates that the more economical bare form is preferred whenever it does not violate the higher ranked faithfulness constraints, even in languages with elaborate functional structure in the nominal domain. In OT terms, the English bare plurals constitute an instance of the 'emergence of the unmarked.'

If Num is not strong enough to introduce a discourse referent, we end up with a definite/indefinite contrast in the singular as well as the plural. An analysis in terms of

weak number has been defended for French (Delfitto & Schroten 1991). We label French a class (vi) language.

(13)   J'ai   lu   *(un)   livre/   *(des)   livres.
          I have   read   *(INDEF_SG)   book/   *(INDEF_PL)   books
          'I read a book/books.'

The singular form *livre* and the plural form *livres* are pronounced in the same way, so they are phonetically indistinguishable. As a consequence of the weak number morphology on the noun, the French bare plural *livres* implies a violation of FDR, whereas its English counterpart *books* in Tableau 2 does not. Interestingly, we do not need additional constraints to account for this case: FDR allows number to introduce a discourse referent, but only if this functional projection is strong. The fact that class (vi) languages have weak number leads to a striking contrast between French and English as far as the possibility of having bare plurals is concerned. French is treated along the same lines as St'át'imcets (cf. 11 above).

The introduction of the constraint FDR completes the set of constraints we need to define our typology. We have seen so far that class (i) languages have the ranking {*Art, *FunctN} >> {FPL, FDEF, FDR}, but probably allow other faithfulness constraints to intervene between *Art and *FunctN. Thus they effectively introduce functional structure (classifiers, case marking, quantification) in the nominal domain (as in 5), without projecting number or using articles, though (cf. 4). Examples are Tai, Japanese and Mandarin Chinese. In the grammar of class (ii)-(vi) languages, various faithfulness constraints are ranked above the markedness constraints *Art and *FunctN, and more and more functional structure is created at the expense of the use of bare nominals. The result is summed up in Table 1.

Table 1: A typology of bare nominals, number morphology and articles

class	ranking	characteristics	example
(i)	{*ART, *FUNCTN} >>{FPL, FDEF, FDR}	no number, no articles	Chinese
(ii)	FPL >> {*ART, *FUNCTN} >> FDEF, FDR	sg/pl distinction, no articles	Georgian
(iii)	{FDEF, FPL} >> {*ART, *FUNCTN} >> FDR	sg/pl distinction, definite/bare contrast	Hebrew
(iv)	{FDR, FPL} >> {*ART, *FUNCTN} >> FDEF	no def/indef; no bare nominals (weak Num)	St'át'imcets
(v)	{FDR, FDEF, FPL} >> {*ART, *FUNCTN}	def/indef contrast, bare plurals OK (strong Num)	English, Dutch, etc.
(vi)	{FDR, FDEF, FPL} >> {*ART, *FUNCTN}	def/indef contrast, no bare nom. (weak Num)	French

In Class (i)-(iii) languages we find bare (singular) nominals in regular argument position; languages in class (iv)-(vi) exclude bare singulars. Whether bare plurals are allowed in these languages depends on the strength of number morphology (weak in St'át'imcets and French, strong in English).

Note that the faithfulness constraints do not interact with each other, so their mutual ranking above/below the markedness constraints *Art and *FunctN is irrelevant. There might be some interaction not visible from this table in that most languages that rank FDEF and/or FDR above *Art also rank FPL above *FunctN. One possible explanation is that higher levels of projection imply lower levels, so a full DP also has a NumP and an NP, and a NumP also has an NP. Corbett (2000: 278, 279) points out that some languages exemplify number distinctions for definites only (Kambera, Basque). In Kambera and Basque, the higher position of D (where the definite article lives) then drives the lower position of Num (where the singular/plural distinction is made) to be filled as well, but bare nominals unmarked for definiteness would not have to convey number information. In fact, this might imply a low ranking for FPL in these languages after all. We leave an elaboration of this idea for future work.

With this caveat, Table 1 represents a full factorial typology in the sense that all possible rankings of the constraints are illustrated, and we do not expect other grammars to be possible for natural language. Accordingly, we claim that the typology captures major classes of number morphology and article use in natural language. The only way to elaborate the classification is to add more constraints, that establish more fine-grained distinctions in the nominal domain (cf. our discussion of class i languages above). Such constraints are also needed to account for languages that have a morphological singular (but use the unmarked form for plural reference), or languages that have an indefinite, but not a definite article. We will leave this possibility for future research. The typology indicates that several distinctions are relevant, and their interaction in the grammar determines the distribution of bare nominals in the language. As a result, bare nominals in one language (belonging to one class) need not have the same range of meanings as bare nominals in another language (belonging to some other class). With the OT typology in place, it is possible to develop an account of the cross-linguistic semantics of bare nominals in terms of blocking.

# 3   Cross-linguistic semantics of bare nominals

The OT typology developed in section 2 allows us to recover the range of interpretations of a bare nominal in a language. The faithfulness constraints FPL, FDEF, and FDR mirror the correspondence rules developed in Section 1. Because of weak morphology and the high ranking of these constraints in (iv) languages like St'át'imcets and class (vi) languages such as French, no bare nominals are generated in regular argument position in these languages, as illustrated in (10, 11, 13). Thus the syntax-

semantics interface of these languages is fully explicit: syntactic projections and semantic interpretations are in perfect correspondence. But what about those languages in which the faithfulness constraints are outranked by *FunctN and *Art? These markedness constraints are operative in the syntax, but not in the semantics. No correspondence rules are defined for bare nominals, so what fixes their interpretation? We propose that bare nominals in class (i), (ii), (iii) and (v) languages get their interpretation by blocking.

3.1   Blocking interpretations

If we go up from the bottom of Table 1, we find bare plurals, but not bare singulars in regular argument position in class (v) languages (English, Germanic). We know that bare plurals in English have a non-definite or generic (plural) meaning. The plural meaning of the bare plural is directly derived from the plural morphology, under the assumption that functional structure is interpreted in terms of the correspondence rules defined in Section 1. The generic reading of English bare plurals is accounted for in this model in Farkas and de Swart (2007). In order to account for the non-definite meaning of bare plurals, we need to adopt a bidirectional view. In the grammar of English, the constraint FDEF is ranked above *Art. This ranking implies marking of determined reference by means of a definite article. Under a strong bidirectional analysis, the definite meaning of English bare nominals is blocked by the definite DP, as illustrated in Figure 1. In this diagram, the two possible forms are at the left-hand side, the two possible meanings at the top. This gives four possible form-meaning pairs, represented by the smaller circles. The arrows represent preferences between the form-meaning pairs, as defined by the constraints. The hands represent the bidirectionally optimal pairs.

Figure 1: Strong bidirectional optimization (English bare plurals)

	non-definite	definite
bare plural	∘ ✌	∘
	↑	↓
definite plural	∘ →	∘ ✌
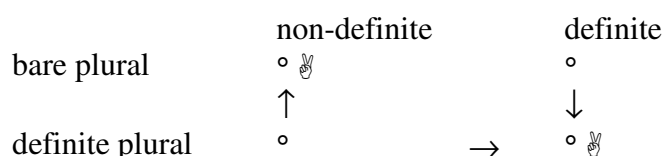
A definite meaning input (determined reference) requires a definite form to satisfy FDEF (down-arrow in last column). A definite form contributes determined reference because of the correspondence rule (right arrow in bottom row). A non-definite meaning input does not lead to a preference for a definite form, and *Art prefers a bare form, so we find an uparrow in the left column. The bare plural itself does not have a preference for a definite or a non-definite interpretation (no arrow in top row), but given that the definite meaning is associated with another form, it is left with a non-definite interpretation only. Under the strong bidirectional analysis in Figure 1, bare plurals always get a non-definite, plural interpretation.

We can extend this argumentation to bare singulars, in languages where they occur. Class (iii) languages such as Hebrew exemplify a contrast between definite and bare nominals in the singular as well as the plural. In both cases, the bare form gets a non-definite interpretation under strong bidirectional optimization, along the lines scketched in Figure 1.

The range of interpretations of a bare singular in class (i) and (ii) languages is wider than those in class (iii) languages, because there is no competition with a definite form. In the absence of a definite article, bare forms can have both definite and non-definite interpretations according to the grammar. The combination of the OT syntax with the correspondence rules thus captures the data from Mandarin Chinese and Georgian (examples 4 and 5 in Sections 1 and 2 respectively).

3.2   Role of argument structure

There is one wrinkle in the syntax-semantics interface set up so far. Most of the work in dynamic semantics focuses on the role of determiners as the linguistic expression that introduces a discourse referent. Farkas and de Swart (2003) assigned plural morphology a dynamic role, which widened the perspective to include class (v) languages. However, we need to go further to account for the class (i)-(iii) languages. The examples in (3), (4) and (5) above, and discussions in Dayal (1999, 2004, 2007) indicate that bare singular nominals in these languages have full discourse referential power. However, under the sparse syntactic assumptions made in this paper, there is nothing in the functional structure of such nominals that introduces a discourse referent. So how do bare singulars get to function in regular argument position, where they must be able to license discourse referents? We assume the general interpretation rule ARG as a way to connect verbs and different nominal projections in regular argument position.

♦   ARG: Parse an XP in argument position as a discourse referent
              (where X = N, Num or D)

ARG relates the presence of some nominal projection (an NP, NumP or DP) in regular argument position to a semantic representation involving a discourse referent. A nominal occurring in a regular argument position instantiates the thematic argument of the predicate by a discourse referent (Farkas and de Swart 2003). Other constructions such as incorporation (Farkas and de Swart 2003) or predication (de Swart, Zwarts and Winter 2007) involve special modes of composition, but these are always morphologically or syntactically 'flagged' as being special. In regular argument position we see the default case of the predicate combining with a discourse referential nominal. The semantic constraint ARG encodes this mode of composition as the unmarked case. All nominals in regular argument position introduce discourse referents according to ARG. Class (i) to (iii) languages just don't reflect the discourse referential status of the nominal argument in the functional projection of the nominal, but use an unmarked (bare) form, as a result of the high ranking of *Art. The interaction of ARG with the

language-specific OT syntax accounts for the observation that bare nominals in class (i) to (iii) languages have full discourse referential status, in line with the claims made by Dayal (1999, 2004, 2007), and others. This completes our account of the syntax-semantics interface in class (i)-(iii) languages.

Notice the following important implication of the appeal to ARG. If an XP is *not* in an argument position, then it does not fall under ARG and, as a consequence, no discourse referent is parsed in that case. In other words, we allow for non-referentiality in non-argument positions. This immediately has consequences for the way noun phrases are expressed in the non-argument positions. As already observed by Longobardi (1994) and others, the usual requirements for referential marking of nominals in type (iv)-(vi) languages often does not hold for nominals in non-argument positions, like predicates or vocatives. (14) and (15) illustrate this for a class (iv) language like St'át'imcets and a class (vi) language like French:

(14)   kúkwpi7   kw   s-Rose
         chief   DET nom-Rose
      'Rose is a chief.'

(15)   Il   est   médecin
        he   is   doctor
      'He is a doctor.'

ARG is not operative in predicative contexts, so no discourse referent is present for these nominals. Therefore the crucial constraint FDR does not apply (cf. de Swart and Zwarts 2008 for an analysis of the special meaning effects arising in these contexts in bidirectional OT). Interestingly, languages differ substantially in how they treat nominals in non-argument positions. English, for instance, does not drop the article of a predicate nominal as easily as many other European languages (de Swart, Winter and Zwarts 2007). There is even more variation in the domain of incorporation (Farkas and de Swart 2003). One interesting but difficult question is how to treat the cross-linguistic variation that we find in such positions, even within the global types that we distinguish in de Swart and Zwarts (2007). This requires a solid characterization of the notion of (non-)argument position and the semantic representation of various non-argument constructions, as well as a view on how finer-grained patterns of variation (with lexical and constructional aspects) can be derived in Optimality Theory.

4   Conclusion

This paper develops a typology of bare nominals, article use and singular/plural morphology in bidirectional Optimality Theory. In this framework, constraints are universal, but they are soft, and can be violated in order to satisfy a more important constraint. The markedness constraints, driving towards minimal functional structure,

and favoring bare nominals conflict with faithfulness constraints driving the expression of cognitive distinctions in atomic/sum reference and discourse referential status. The ranking of constraints is language specific, and the factorial typology produces a range of possible grammars, that are exemplified by various languages. The cross-linguistic variation in the semantics of bare nominals is derived from the typology: certain meanings can be blocked for bare nominals, because they are realized by more specific forms, but this depends on the grammar of the language at hand. In further research, we intend to test the empirical adequacy of the typology against a wider range of languages, and elaborate the set of constraints.

References

Anderson, Stephen and Ed Keenan (1985) "Deixis", in T. Shopen (ed.). *Language typology and syntactic description*, Cambridge: CUP, 259–308.

Borer, Hagit (2004) *Structuring Sense*, Oxford: Oxford University Press.

Chierchia, Gennaro (1998a) "Reference to kinds across languages", *Natural Language Semantics* **6**, 339–405.

Chierchia, Gennaro (1998b) "Plurality of mass nouns and the notion of 'semantic parameter'", in S. Rothstein (ed.). *Events and Grammar*, Dordrecht: Kluwer, 53–103.

Corbett, Greville (2000) *Number*, Cambridge: Cambridge University Press.

Dayal, Veneeta (1999) "Bare NPs, reference to kinds and incorporation", in *Proceedings of SALT 9*, Ithaca: Cornell University Press.

Dayal, Veneeta (2007) *Hindi pseudo-Incorporation*, Ms. Rutgers University.

Dayal, Veneeta, (2004) "Number marking and indefinites in kind terms", *Linguistics and Philosophy* **27**, 393–450.

Delfitto, Denis and Jan Schroten (1991) "Bare plurals and the number affix in DP", *Probus*, **3**, 155–185.

Doron, Edit (2004) "Bare singular reference to kinds", in R.B. Young and Y. Zhou (eds.) *Proceedings of SALT 13*.

Farkas, Donka (2002) "Specificity distinctions", *Journal of Semantics* **19**, 1–31.

Farkas, Donka (2006) "The unmarked determiner", L. Tasmowski and S. Vogeleer (eds.) *Non-Definiteness and Plurality,* Amsterdam: Benjamins, 81–106.

Farkas, Donka and Henriëtte de Swart (2003) *The semantics of incorporation*, Stanford: CSLI Publications.

Farkas, Donka and Henriëtte de Swart (2007) "Article choice in plural generics", *Lingua* **117**, 1657–1676.

Farkas, Donka and Henriëtte de Swart (2008) "Stability and variation in article choice: generic and episodic contexts", in A. Giannakidou and M. Rathert (eds.) *Quantification, nominalization and definiteness,* Oxford: OUP (in press).

Greenberg, Joseph (1963) "Some universals of grammar with particular reference to the order of meaningful elements", in J. Greenberg (ed.). *Universals of Language*, Cambridge: MIT Press, 73–113.

Greenberg, Joseph, Charles Ferguson and Edith Moravcsik (1978) *Universals of human language*, Stanford CA: Stanford University Press.

Harris, Alice (1981) *Georgian syntax: a study in relational grammar*, Cambridge: Cambridge University Press.

Heim, Irene (1982) *The semantics of definite and indefinite noun phrases in English,* PhD thesis, University of Massachusetts at Amherst.

Higginbotham, James (1985) "On Semantics", *Linguistic Inquiry* **16**, 547–593.

Kamp, Hans (1981) "A theory of truth and semantics representation", in J. Groenendijk, T. Janssen & M. Stokhof (eds.) *Truth, representation and information*, Dordrecht: Foris, 277–322.

Kamp, Hans and Uwe Reyle (1993) *From discourse to logic*, Dordrecht: Kluwer Academic Press.

Kamp, Hans and Jan van Eijck (1997) "Representing discourse in context", in J. van Benthem and A. ter Meulen (eds.) *Handbook of Logic and Linguistics*, Amsterdam: Elsevier, 179–237.

Krifka, Manfred (1995) "Common nouns: a contrastive analysis of Chinese and English", in G. Carlson and J. Pelletier (eds.) *The generic book*, Chicago: University of Chicago Press, 398–411.

Longobardi, Giuseppe (1994) "Reference and proper names", *Linguistic Inquiry* **25**, 609–665.

Matthewson, Lisa (1998) *Determiner systems and quantificational strategies: Evidence from Salish*, the Hague: Holland Academic Graphics.

Rijkhoff, Jan (2002) *The noun phrase*, Oxford: Oxford University Press.

Swart, Henriëtte de and Joost Zwarts (2008) "Less form – more meaning: why bare singulars are special", *Lingua* (in press).

Swart, Henriëtte de, Yoad Winter and Joost Zwarts, (2007) "Bare nominals and reference to capacities", *Natural Language and Linguistic Theory* **25**, 195–222.