

The Distribution of two Indefinite Articles – The Case of Uzbek

Klaus von Heusinger & Udo Klein

University of Stuttgart & University of Bielefeld

To appear in: C. Ebert & S. Hinterwimmer (eds.). *Different Kinds of Specificity across Languages*. Dordrecht: Springer (Studies in Linguistics and Philosophy=

Abstract

The aim of this paper is to show that Uzbek has the crosslinguistically rather unusual property of using two indefinite articles, namely the expressions *bir* and *bitta*. From a synchronic perspective the two indefinite articles can be shown to compete (at least in certain environments), and from a diachronic perspective there is evidence that *bitta* significantly expands to the expense of *bir*. The in-depth investigation of corpus data and data from an online-questionnaire contributes to the discussion of the diachronic development of indefinite articles in general.

1. Introduction*

Uzbek is one of the rare languages that make use of two indefinite articles to indicate indefinite noun phrases, as in (1a-b).

* We would like to thank our informant Dildora Niyazmetova for her patience in answering our questions, for annotating the Uzbek corpus which we used, and for her help with the questionnaire, as well as Jaklin Kornfilt, Elena Skribnik and Dolgor Guntsetseg for discussion of an earlier version of this paper. We are also very grateful for the very helpful comments of the editors, Cornelia Ebert and Stefan Hinterwimmer and their continuous efforts editing this volume. All remaining errors are, of course, our own. The present work has been carried out as part of the project C2 “Case and referential properties” of the collaborative research centre SFB 732 “Incremental specification in Context” of the German Science Foundation, whose financial support we gratefully acknowledge.

- (1) a. Peter *bir* qiz-ni kusat-ayap-ti
Peter a girl-ACC observes-PRS.PROG-3.SG
‘Peter observes a girl.’
- b. Peter *bitta* qiz-ni kusat-ayap-ti
Peter a girl-ACC observes-PRS. PROG-3.SG
‘Peter observes a girl.’

Since the status of *bir* as an indefinite article is already controversial, e.g. Bodgrogligeti (2001, pg. 55) claims in his grammar that there are no definite or indefinite articles in Uzbek, and that the expression *bir* only expresses the numeral ‘one’, it is not surprising that there is no discussion of a second indefinite article, the distribution and the potential semantic contrasts. Contrary to these claims, we show that Uzbek has not just *one* but *two* indefinite articles, namely *bir* and *bitta*. Uzbek thus shares this crosslinguistically rather unusual property with languages such as Lakhota, which distinguishes between two “quasi-indefinite determiners” (Lyons 1999), Moroccan Arabic, which employs a specific and a non-specific indefinite article (Harrell 1962, Fassi-Fehri 2006), and Maori, a more well-known example of a language where we also find two indefinite articles (Bauer 1993). The close investigation of the distribution of the two indefinite articles in Uzbek does not only provide an interesting insight into the semantic and pragmatic conditions regulating this distribution, but it also allows us to address some more general questions with respect to indefiniteness and indefinite articles, such as: (i) can we describe indefiniteness as one semantic/pragmatic category or as a familiarity concept of different ones (cf. the discussion of the category of definiteness in Lyons 1999 and others); (ii) how is the domain of indefiniteness divided by two or more markers; (iii) what can we learn from the diachronic perspective on the grammaticalization of indefinite articles, and (iv) what methods can we apply to investigate the distribution and the semantic function of indefinite articles.

In section 2 we provide a brief overview of article systems with two indefinite articles and discuss Heine’s grammaticalization path for indefinite articles that developed from the numeral for ‘one’. We formulate some questions that concern the generality of such assumptions in systems with more than one indefinite article. Uzbek constitutes a very special case since both its indefinite articles derived from two forms of the numeral for ‘one’. In section 3, we provide some information about the structure of nominals in Uzbek and we

show that *bir* and *bitta* are indeed used as indefinite articles. We discuss some contrasts of the two indefinite articles with respect to scope and specificity.

In section 4 we investigate the diachronic development of the two articles by comparing their occurrences in two novels, the first one published in 1926 and the second one published in 2001, and show that while *bir* has reached the last development stage of Heine's scale already by 1926, the use of *bitta* is a more recent development (at least in written Uzbek). There is, however, an interesting additional observation from the small corpus we used: The frequency of indefinite noun phrases with *bir* in the text from 1926 is about 10% of all indefinite noun phrases and the frequency of indefinite noun phrases with *bir* and with *bitta* is about 4% of all indefinite noun phrases in the text from 2001, which – if it could be generalized – would indicate a decrease of marking indefiniteness with indefinite articles. The second important conclusion from this corpus study is that the use of *bitta* appears to have spread to the expense of *bir* not just in written but also in colloquial Uzbek. Due to the inherent limitations of corpus studies, some important questions could not be answered by this methodology. In section 5 we present the results of a web-based questionnaire, which was designed to elicit acceptability judgements in order to address some of these questions. First, the fact that certain usages of *bitta* are not attested in our corpus does not show that *bitta* cannot be used that way (absence of evidence for a particular usage of *bitta* is, of course, not evidence for the absence of this usage). Acceptability judgements can help settle this issue. Secondly, we wanted to know whether, and if so in which contexts of use, in modern spoken Uzbek the use of *bitta* is judged better than the use of *bir*, because if this is the case then this would corroborate the hypothesis that *bitta* has spread to the expense of *bir*. The main result is that when it is used as an indefinite article introducing human referents, *bitta* is at least as good as *bir* (except in predicative constructions), and indeed better if used to introduce referents whose identity neither the speaker nor the hearer knows or cares about. These results also contribute to the discussion of how to divide the domain of indefiniteness¹ and whether the diachronic evolution of indefinite articles follows such a division.

¹ See Haspelmath (1997) for indefinite pronouns in general, Yanovich (2012) for indefinite pronouns in Russian, and Ebert & Hinterwimmer (2012) for specificity markers in German.

2. Indefinite articles

2.1 Article systems with more than one indefinite article

Dryer (2011) reports from a survey of more than 500 languages with respect to their article systems (percentages added by us):

Table 1 Types of article systems (Dryer 2011)

type of article system	instances	percentages
Indefinite word distinct from numeral for 'one'	102	19%
Numeral for 'one' is used as indefinite article	112	21%
Indefinite affix on noun	24	4%
No indefinite article but definite article	98	18%
Neither indefinite nor definite	198	37%
total	534	100%

Nearly half the languages in this corpus have indefinite articles, sometimes identical with the numeral 'one', and sometimes also in different forms, e.g. English *a* counts as different from the numeral *one*, since diachronic relations are not accounted for. Dryer does not distinguish crosslinguistically unusual systems with more than one indefinite article. There are, however, some studies of such systems. There are languages that encode specificity contrasts in the article system like in ChiBemba, a Bantu language, as in (2) (Givón 1973, pg. 102). The VCV-prefix signals referentiality or specificity, and the CV-prefix non-referentiality, both are neutral with respect to definiteness.

- (2) a. umuana aa-fwaayile **ici**-tabo
'The child was looking for **a/the** book.' (referential)
- b. umuana aa-fwaayile **ci**-tabo (cili cyonse)
'The child was looking for **a** book.' (nonreferential)

Blass (1990, pg. 185) reports from two indefinite articles in Sissala, a Niger-Congo language spoken in Burkina Faso, which also signal specificity.²

² It is interesting to compare this with the German contrast between weak and strong definite articles in

- (3) a. ᵁ ᵁ ᵁᵁᵁ ᵁᵁ.
 ‘He went to a market.’ (non-specific)
- b. ᵁ ᵁ ᵁᵁᵁ ᵁᵁ.
 ‘He went to a market.’ (specific)

Similar contrasts are reported for Maori (Bauer 1993, Chung and Ladusaw 2004) and some other languages. Moroccan Arabic provides a three way system of indefinite marking: (i) bare nouns are not marked for specificity; (ii) a specific indefinite article *wahed-l*, composed of the numeral ‘one’ and the definite article, (iii) and a non-specific indefinite article *shi*, derived from the word for ‘thing’, as illustrated in (4a-c) from Fassi-Fehri (2006) (see Brustad 2000: 26-31 for other Arabic dialects), and summarized in table 2:

- (4) a. meryem bgha-t te-t-zewwej b-muhami wa-layenni waldii-haa
 Maryam wanted-F to-F-marry with-lawyer but parents-her
 ma bghaw-eh-sh / wa-layenni ma lqa-t-u-sh
 not wanted-him-neg / but not met-her-him-NEG
 ‘Maryam wanted to marry a lawyer but her parents don’t like him/ but she has not met one yet.’
- b. meryem bgha-t te-t-zewwej b-**wahed** r-rajel wa-layenni ma lqa-t-u-sh
 Maryam wanted-F to-f-marry with-**one** **the**-man but not met-her-him-NEG
 ‘Maryam wanted to marry a (specific) man but she hasn’t found him (* one).’
- c. meryem bgha-t te-t-zewwej b-**shi** rajel wa-layenni ma lqa-t-u-sh
 Maryam wanted-F to-F-marry with-**some** man but not met-her-him-NEG
 ‘Maryam wanted to marry a (non-specific) man but she hasn’t found one.’

(i)-(ii). The weak (non-referential form) merges with the preposition *in* ‘to’ to the form *ins*, while the strong (referential) form *in das* does not allow such merging (see Schwarz 2009).

- (i) Er geht ins Gefängnis.
 ‘He goes to prison.’
- (ii) Er geht in das Gefängnis.
 ‘He goes to the prison.’

Table 2 Indefiniteness marking in Moroccan (Fassi-Fehri 2006)

Moroccan	specific	non-specific	non-argumental
unmarked	+	+	+
<i>wahel-l</i>	+	-	-
<i>shi</i>	-	+	-

The division of indefinite noun phrases in specific and non-specific subclasses seems to be a natural one (see Farkas 2002, von Heusinger 2011a), but can also be motivated by the meaning of the expressions from which the two indefinite articles are derived: The specific indefinite article *wahel-l* derives from a combination of ‘one’ and the definite article originally providing a kind of partitive reading. The non-specific indefinite article *shi* derives from a bleached form of *shai* ‘thing’, perhaps originally used as negative polarity item or as an enforcer.

2.2 Evolutionary scales for indefinite articles

Givón (1981) proposes an evolutionary scale for the development of the numeral ‘one’ to an indefinite article, as in (5). Quantification describes the semantics of the numeral ‘one’, referentiality / denotation covers the different semantic and pragmatic function of indefinite articles in sentence and discourse and genericity indicates that the indefinite article is used as a generalized functional element for noun phrases.

(5) quantification -> referentiality/denotation -> genericity.

This scale clearly predicts a contrast between quantificational (or cardinal) readings, referential readings that introduce a discourse referent, and non-argumental readings (as in predicational constructions). However, it is not clear why and how the grammaticalization process proceeds and what the intermediate contexts are. Heine (1997, p. 66-82) elaborates this three stage model to a more fine-grained five stage model, as in (6), in order to account for the diachronic evolution of an indefinite article from a numeral. In the first stage, “there is an item for ‘one’ which functions exclusively as a numeral” (Heine 1997, p. 72). Such an item reaches the second stage in its development into an indefinite article when it “introduces

a new participant presumed to be unknown to the hearer and this participant is taken up as definite in subsequent discourse” (72). In the third stage, the use of this item “is extended typically to any participant in discourse known to the speaker but presumed to be unknown to the hearer, irrespective of whether or not the participant concerned is expected to be taken up in subsequent discourse” (72f). The fourth stage is reached if the item can “be used when a participant is introduced whose referential identity neither the hearer nor the speaker knows or cares to know” so that the article is “no longer confined to marking specific reference at this stage” (73). In the fifth stage, “the article can be expected to occur on all types of nouns, even if there may remain a number of exceptions.” Heine emphasizes that “[t]he various stages must not be viewed as discrete entities; rather the evolution from stage I to V is continuous and involves overlaps of various kinds.”

- (6) I: numeral -> II: presentative marker -> III specific marker
-> IV non-specific marker -> V: generalized article

Heine’s five stage model can be viewed as an elaboration of the three stage model proposed in Givón (1981). Heine’s first stage corresponds to Givón’s quantification category, Heine’s second, third and fourth stages can be viewed as a subclassification of Givón’s referentiality category, while Heine’s fifth stage corresponds to Givón’s genericity category. In annotating our corpus, we used Heine’s system, see section 5.³

3. The system of indefinite articles in Uzbek

Bodrogligeti (2001, pg. 55) claims that in Uzbek “[t]here is no special morphological or grammatical means specifically designed to distinguish between definite and indefinite nouns: No definite or indefinite articles”. The (in)definiteness of noun phrases is determined by the position in the sentence and the pragmatic context (as in other article-free languages).

³ Schroeder (2006) suggests a modification of Heine's model. He assumes a stage 1 “emergent indefinite article” that is specific, stage 2 “pragmatic indefinite article”, stage 3 “referential indefinite article” and stage 4 “extended indefinite article”, which covers predicational uses. He thus differs from Heine in exchanging the pragmatic and the specific indefinite article in their order. Heine assumes with Givón (cf. Wright and Givón 1987). that the first use of a numeral in non-cardinal contexts signals pragmatic prominence and only in a second step semantic specificity, while Schroeder assumes that in the first step it signals specificity, and only in a second discourse salience. See von Heusinger (2011b) for a similar discussion for German indefinite *dies* and indefinite *so'n*. See also the extensive study of indefinite pronouns by Haspelmath (1997).

The definiteness in (7) must be derived from the topic position of *qiz* ‘girl’ and the pragmatic context for *ampir* ‘old woman’. Demonstrative pronouns can be used to unambiguously mark definite expressions, as in (8), but they are not definite articles since they are ungrammatical in associative uses, as in (9), where the possessive marker *-i* marks the bridging relation.⁴

- (7) Qiz sevin-ib kampir old-i-ga kir-ib-di
 girl happy-EVID old-woman to-POSS-DAT enter-EVID-3.SG

‘The girl was happy and entered towards the old woman.’

- (8) Bu/shu odam ko’p narsa bil-a-di
 this / that man much thing know-PRS.PROG-3SG

‘This / that man knows much.’

- (9) Men bitta kitob o’qi-di-m, *bu/shu muallif / muallif-i juda-yam mashhur
 1SG a book read-PRF-1.SG *this/that author / author-POSS very-SPRL famous

‘I read a book, the (=its) author is famous.’

Indefinite noun phrases are generally unmarked, as in (10) and (11), but they can be marked by indefinite pronouns or by numerals as in (12) or by the expressions *bir* and *bitta*, as in (1), repeated as (13).

- (10) O’rmon-da turli daraht-lar o’s-a-di
 forest-DAT different tree-POL grow-FUT-3.Sg

‘Different trees grow in a forest.’

- (11) Farhod olma arch-ayap-ti
 Farhod apple peel-PRS.PROG-3.SG

‘Farhod peels apple / an apple.’

⁴ We use the following abbreviations: ABL – ablative, ACC – accusative, AGR – agreement, CLF – classifier, CVB – converg, COND – conditional, DAT – dative, DEM – demonstrative, EVID – evidential, F – feminine, FUT – future, GEN – genitive, IMP – imperative, LOC – locative, NEG – negation, PL – plural, POSS – possessive, PRF – perfect, PROG – progressive, PRS – present, PTCP – participle, SG – singular, SPEC – specifier (in contrast to a classifier), SPRL - superlative

(12) U bir dona olma e-di
 He/she one CLF apple eat-3.SG
 ‘He/she eats apple / an apple.’

(13) a. Peter *bir* qiz-ni kusat-ayap-ti
 Peter a girl-ACC observes-PRS.PROG-3.SG
 ‘Peter observes a girl.’
 b. Peter *bitta* qiz-ni kusat-ayap-ti
 Peter a girl-ACC observes-PRS-PROG.3.SG
 ‘Peter observes a girl.’

We agree with Bodrogligeti’s claim that there is no definite article, but we disagree that there is no indefinite article. We argue in this section that Uzbek has not only one article, but two, namely *bir* and *bitta*. Both of them developed from the numeral *bir* ‘one’. Since this is an instance of a general pattern of diachronic development of indefinite articles from numerals, we follow the five-stage model developed in Heine (1997) to account for this development, and turn to the basic ingredients for the development of these indefinite articles, namely the numeral *bir* ‘one’ and the classifier system of Uzbek. With these important preliminaries clarified, we turn to the two items *bir* and *bitta*, and indicate what stage they have reached in their development into indefinite articles.

3.1 Numerals and classifiers in Uzbek

Since in Uzbek, the indefinite articles have their root in the numeral *bir* ‘one’, in this subsection we will briefly present the main property of numerals in Uzbek (which sets Uzbek apart from other Turkic languages), namely that they require what Beckwith (1998) calls a specifier if they occur with a count noun. We will use the terminology used in Beckwith (1998), since this is the most detailed and theoretically informed paper on Uzbek classifiers that we are aware of. At the end of this subsection we will provide what we think are the corresponding notions in Aikhenvalds terminology.

Beckwith (1998) distinguishes two main classes of specifiers (i.e. lexical items which must occur with numerals plus count nouns) in Uzbek, namely measures and classifiers.

- (18) a. ikki-ta hona
 two-SPEC room
 ‘two rooms’
- b. *ikki hona
 two room
 Int.: ‘two rooms’

As the examples above show, expressions containing a numeral and a count noun are ungrammatical if they do not contain a measure as in (16), a classifier as in (17), or the nonclassifying specifier suffix *-ta* as in (18).

The only exception to this generalisation is the numeral *bir* ‘one’, which may also occur without a specifier: “some speakers in Uzbekistan, especially in the cities, generally drop the suffix with the numeral *bir* ‘one’” (Beckwith 1998, pg. 127). That is, with the numeral *bir* it is not necessary to use a specifier (either a measure, a classifier or the nonclassifying specifier suffix *-ta*), so that *bir* (19a) can be used instead of *bitta* (19b), which as suggested by Beckwith is the result of the assimilation of the numeral *bir* to the nonclassifying specifier suffix *-ta*.

- (19) a. bir qiz
 one girl
 ‘one/a girl’
- b. bitta qiz
 one:SPEC girl
 ‘one/a girl’

According to Beckwith, this “may have developed due to influence from the use of *bir* as an indefinite article, in which usage no specifier occurs” (127). Note that the way in which Beckwith formulates this exception to the rule that numerals require a specifier implies that *bir* can occur without a specifier not only when it is used as an indefinite article, but also when it is used as a numeral. In the next subsection we will illustrate the distribution of *bir* as an indefinite article, in order to show which stage in the development it has reached.⁵

⁵ A brief terminological note: Beckwith’s class of measures corresponds to Aikhenvald (2000)’s class of

3.2 The indefinite article *bir*

In this subsection we show that, contrary to claims made in the literature according to which Uzbek does not have indefinite articles, the numeral *bir* ‘one’ can indeed be used as an indefinite article, which was already noted by von Gabain (1945, pg. 92): “*Bir* ‘eins’ dient auch als unbestimmter Artikel: *bir jigiti* ‘ein junger Mann’.” (*Bir* “one” also serves as indefinite article: *bir jigiti* “a young man”.) In section 4 we provide more written material that shows that *bir* was used as an indefinite article already in 1926.

In example (20) *bir* is used to introduce a new participant which (i) is presumed to be unknown to the hearer and which (ii) is the subject of further specification in the subsequent discourse.

- (20) *Bir zamon-lar bir schoh bo'l-ib o't-gan ekan...*
 A time-PL a(=one) king be-CVB be-PTCP apparently
 ‘Once upon a time there was a king. . .’

According to Heine (1997, pg. 66-82), this is the second stage in the development of an indefinite article from a numeral. Sentence (21) below, taken from Tohir Malik’s novel *Shaytanat*, illustrates that *bir* can also be used to introduce a new referent without picking up this referent in the following discourse (stage 3); sentence (22) illustrates that *bir* can be used without introducing a specific referent (stage 4); and sentence (23) shows that *bir* can be used in predicative constructions (stage 5). The following sentences are taken from two Uzbek novels (see next section).

- (21) *Eshpo'latov ichkari bir erkak-ni ko'r-gan. hayron*
Eshpolatov inside-LOC a man-ACC see-PRF wonder
bo'l-ib zavod-ga bor-gan.
be-CVB workshop-DAT walk-PRF
 ‘Eshpolatov saw a man inside. Bewildered he walked into the workshop.’

mensural classifiers, whereas Beckwith’s classifiers correspond to Aikhenvalds’s notion of sortal classifier. See Aikhenvald (2000, pg. 114-120) for criteria distinguishing these types of classifiers.

- (22) Uch-av-lar-i ham bir narsa-ning xayol-i-ni
 three-together-PL-AGR also a thing-GEN think-AGR-ACC
 sur-gan-dek ko`r-i-nar-di-lar.
 do-PTCP-as.if see-AGR-PROG-3-PL

‘The three looked as if they were thinking about something.’

- (23) Hurmat-i zo`r bo`l-sa boshqa-lar-ga zo`r, sen bilan
 honor-AGR strong be-COND other-PL-DAT strong, 2.SG with
 men-ga u bor-yo`g`i bir oshna.
 1.SG-DAT 3.SG be-just a friend

‘If he is very honorable, then only for the others. For you and for me he is just a friend.’

We therefore conclude that the Uzbek numeral *bir* has reached the fifth stage of the development into an indefinite article. In the next subsection we turn to the development of *bitta* into an indefinite article.

3.3 The indefinite article *bitta*

Uzbek differs from other Turkic languages, e.g. Turkish which also uses the word *bir* as an indefinite article (Kornfilt 1997, pg. 106), in that in addition to *bir* it has developed a second indefinite article, namely *bitta*. The form *bitta* derives from *bir* + *ta*, i.e from the numeral ‘one’ and the nonclassifying specifier suffix *-ta*.⁶

The following sentences, found again in Tohir Malik’s novel *Shaytanat*, illustrate that it has reached at least the fourth stage of development into an indefinite article. In (24) it is used to introduce a new referent which is then picked up again in the following discourse (stage 2), in (25) it is used to introduce a new referent which is not picked up again (stage 3), and in (26) it is used without introducing a specific referent (stage 4). In the contemporary novel which we analysed we found no use of *bitta* in a predicative construction (see next section):

⁶ Bodrogligeti (2003, pg. 456): “The suffix *-ta* attached to cardinal numbers creates a complete set of numbers used to count objects by units. They are followed by nouns in the singular.” Von Gabain (1945, pg. 91): “+ta, t̪ (<pers.). Einfache Zahlwörter werden häufig durch dies Formans erweitert, das ursprünglich ein Zahlwort für Sachen und Personen war, und das im Özb. [= Uzbek] nicht vor Mengenbezeichnungen gebraucht wird.” Boeschoten 1998, pg. 363): “The cardinal numbers one to nine are *bir, ikki, üç, tört* [...]’. Optionally *-ta* may be added, in particular to lower numerals, e.g. *bittá, ikkitá, [...]*.”

(24) Ammo bitta shart-i-m bo'l-a-di: singl-i-m-ni
 but a condition-AGR-1.SG be-FUT-3.SG little.sister-AGR-1.SG-ACC
 bun-dan buyon qimor-ga tik-may-san.
 now-ABL from game-DAT put-NEG-2.SG.FUT
 'But there will be a condition; from now onwards you will not bet my little sister.'

(25) Bitta shahar-da ikki-ta imom-domla bo'l-gan ekan. It bilan
 a city-LOC two-SPEC imam-priests be-PRF EVID dog with
 mushuk murosa qil-sa qil-arkan-ki, bu-lar ittifoq
 cat agreement make-COND make-EVID-CONJ DEM-PL agreed
 bo'la ol-ish-mas ekan-lar.
 be take-NOM-NEG EVID-PL
 'In a city there were two imams. Even dogs and cats could agree, but these two couldn't.'

(26) Katta-ng-ga bor-ib ayt: men-ga bitta samolyot
 big-2.SG.AGR-DAT go-CVB say.IMP 1.SG-DAT a plane
 bilan bir million pul kerak.
 with a million money need
 'Go to your boss and say: I need a plane and a/one million.'

3.4 Semantic contrast of indefinite articles

The availability of two indefinite articles may be exploited to mark certain differences between indefinite expressions. As we have seen in section 2.1, some languages, such as Sissala or Moroccan, mark the specific vs. non-specific contrast with the two articles. However, the notion of specificity is too vague and has too many different incarnations (see von Heusinger 2011a), so that we cannot generalize from such singular observations to a more general picture. At least for Uzbek, the picture is not so clear. Our informant reported a certain preference of *bir* with animate and of *bitta* with inanimate nouns. For the classical tests for specificity she also reported certain preferences. In (27a) *bir* signals epistemic non-specificity, while *bitta* in (27b) has a strong preference for a specific reading.

- (27) a. Farhod bir talaba-nı o'p-di
 Farhod a student-ACC kiss-3.SG.
 'Fahrhod kissed a student.' (non-specific)
- b. Farhod bitta talaba-nı o'p-di
 Farhod a student-ACC kiss-3.SG
 'Fahrhod kissed a student.' (specific, non-specific)

We find a similar contrast for indefinite noun phrases under verbs of propositional attitudes. In (28a) there appears to be a strong preference for interpreting the direct object NP non-specifically (i.e. as introducing a referent whose identity the speaker and hearer neither know nor care about), whereas in (28b) the indefinite can but does not have to be so interpreted. The construction (28c) with the numeral and the classifier just gives the cardinal number (see Niyazmetova 2009).

- (28) a. Professor bir talaba-nı tekschir-moq'-chi.
 professor a(=one) student-ACC examine-want-3.SG
 'The professor wants to examine a student.' (non-specific)
- b. Professor bitta talaba-nı tekschir-moq'-chi.
 professor a(=one:SPEC) student-ACC examine-want-3.SG
 'The professor wants to examine a student.' (specific, non-specific, cardinal)
- c. Professor bir nafar talaba-nı tekschir-moq'-chi.
 professor one CLF:HUMAN student-ACC examine-want-3.SG
 'The professor wants to examine one student.' (cardinal)

Even though we were not able to confirm these judgements by other informants, we will assume that the contrast between *bir* and *bitta* signals specificity (for the sake of the argument). The question that arises is whether this contrast reflects their main function of being a specific or a non-specific indefinite article or whether it is an epiphenomenon of a different contrast.⁷ These questions address one of the main issues discussed above: How can indefinite articles divide the domain of indefiniteness into subdomains, and are these

⁷ Compare the study of Keenan & Ebert (1973) on two definite articles in Malagasy and Frisian. They claim that the two articles primarily express the contrast between situational (anaphoric) use and encyclopaedic (unique) use. The articles in both languages also express a secondary use, namely the contrast between a referential and an attributive use.

subdomains “natural” domains or not? In order to investigate these issues we first undertake a diachronic study on the development of the two articles, and secondly we have tested their usages in contemporary Uzbek.

4. Corpus study

To throw some light on the first question, we initially analysed the use of *bir* and *bitta* in the novel *Otgan kunlar* by Abdulla Qodiriy, which was published in 1926, and then we compared this with the use of *bir* and *bitta* in Tohir Malik’s novel *Shaytanat*, published in 2001.

In the novel from 1926, we annotated the first 113 occurrences of *bir* and *bitta* as indefinite articles with respect to the kind of usage and with respect to the animacy of the referent. These 113 occurrences were found among 1080 indefinite noun phrases, i.e. they constitute about 10% of the indefinite noun phrases.⁸ As table 3 shows *bir* has been employed in all four types of usages, from introducing a referent which is picked up again (stage 2) to occurrences in predicative constructions (stage 5). However, in the older novel we did not find *bitta* used as an article (we only found one use of *bitta* as a numeral).

⁸ The percentages of *bir*- and *bitta*-indefinites in both texts are rough approximations: We counted all bare and non-bare noun phrases on the first three pages of both novels and categorized them according to definite and indefinite interpretation. In order to collect enough instances of *bir*- and *bitta*-indefinites we had to analyze the first 85 pages of the 1926 novel and the first 150 pages of the 2001-novel. We then projected the number of all definite and all indefinite to 150 pages and calculated from this the percentage of *bir*- and *bitta*-indefinites, as in table (i) and (ii):

Table (i) Calculation of percentages of *bir*- and *bitta*-indefinites in the 1926-novel

1926 novel	all	def	indef	<i>bir</i>	<i>bitta</i>
first 3 pages	104	64	40		
85 pages	2808	1728	1080	115	0
% of all NPs	100%	61,5%	38,5%	4,1%	0,0%
% of indef NPs	-	-	100%	10,6%	0,0%

Table (ii) Calculation of percentages of *bir*- and *bitta*-indefinites in the 2001-novel

2001 novel	all	def	indef	<i>bir</i>	<i>bitta</i>
first 3 pages	199	109	90		
150 pages	9950	5450	4500	101	75
% of all NPs	100%	54,8%	45,2%	1,0%	0,8%
% of indef NPs	-	-	100%	2,2%	1,7%

Table 3 Use of *bir/bitta* in the novel from 1926

Article	Usage	Occurrences	human	inanimate
<i>bir</i>	2	29	14	15
<i>bir</i>	3	29	16	13
<i>bir</i>	4	41	8	33
<i>bir</i>	5	14	11	3
<i>total</i>		113	49 (43%)	64 (57%)
<i>bitta</i>	2	0	0	0
<i>bitta</i>	3	0	0	0
<i>bitta</i>	4	0	0	0
<i>bitta</i>	5	0	0	0
<i>total</i>		0	0	0

There are two (logically) possible reasons for why *bitta* is lacking in this novel. Either, *bitta* was not yet in use as an indefinite article in 1926, or it was already in use, but not yet in the written register. Assuming that the written register lags behind the spoken register, it is plausible to conclude that *bitta* has developed into an indefinite article after *bir*.

The lack of *bitta* in the novel published in 1926 stands in sharp contrast to the presence of *bitta* in the novel from 2001, in which we annotated the first 125 occurrences of *bir* and *bitta* as indefinite articles out of 4500 indefinite noun phrases. This means that the overall usage of *bir* and *bitta* together is not more than 2,7% or one fourth of the percentage of *bir* in the 1926 novel. As illustrated in table 4, approximately two thirds (81) were occurrences of *bir*, and approximately one third (44) were occurrences of *bitta*.

Table 4 Use of *bir/bitta* in the novel from 2001

Article	Usage	Occurrences	human	inanimate
<i>bir</i>	2	16	2	14
<i>bir</i>	3	40	8	32
<i>bir</i>	4	16	3	13
<i>bir</i>	5	9	2	7
<i>total</i>		81	15 (12%)	66 (53%)
<i>bitta</i>	2	12	2	10
<i>bitta</i>	3	20	7	13
<i>bitta</i>	4	12	4	8
<i>bitta</i>	5	0	0	0
<i>total</i>		44	13 (10%)	31 (25%)

Given that in the first novel the lack of *bitta* may have been due to differences in register between written and colloquial language, we also annotated whether the indefinite articles *bir/bitta* occurred in the author's narrative, which is likely to be closer to the written register, or whether it occurs in direct speech, which is likely to be closer to the colloquial register. As illustrated in table 5, out of 66 occurrences of indefinite articles in the speaker's narrative, 62 were occurrences of *bir*, and only 4 were occurrences of *bitta*. From this we can conclude that there is a strong preference for *bir* in the speaker's narrative. On the other hand, out of 59 occurrences of indefinite articles in direct speech, only 19 were occurrences of *bir*, compared to 40 occurrences of *bitta*. This indicates that in direct speech there is a preference for using *bitta*.

Table 5 Use of *bir/bitta* in the novel from 2001, annotated by narrative versus direct speech

Article	Usage	Occurrences	Narrative	Direct speech
bir	2	16	13	3
bir	3	40	33	7
bir	4	16	11	5
bir	5	9	5	4
total		81	62 (50%)	19 (15%)
bitta	2	12	1	11
bitta	3	20	0	20
bitta	4	12	3	9
bitta	5	0	0	0
total		44	4 (3%)	40 (32%)

Regarding *bir*, we can therefore conclude that its use as an indefinite article had already spread to Heine's fifth stage by 1926, as the examples from the 1926 novel indicate, in which *bir* is used in predicative constructions.

- (29) Xon qizig'a loyiq bir yigit.
khan very worthy a(=one) young.man
'The khan is a very worth young man.'

Regarding *bitta*, we have evidence that in the written register, the use of *bitta* as an indefinite article has increased from the 1926 novel, where we found no occurrence of *bitta* as an indefinite article in the first 113 occurrences of *bir* and *bitta*, to the 2001 novel, where we found 44 occurrences of *bitta* in the first 125 occurrences of *bir* and *bitta* as indefinite articles. Moreover, as the example in (25), repeated below as (29), illustrates, *bitta* may be used to introduce a participant “whose referential identity neither the hearer nor the speaker knows or cares to know”, and has thus reached at least stage 4 of development into an indefinite article.

- (30) Katta-ng-ga bor-ib ayt: men-ga bitta samolyot
 big-2.SG.AGR-DAT go-CVB say.IMP 1.SG-DAT a plane
 bilan bir million pul kerak.
 with a million money need
 ‘Go to your boss and say: I need a plane and a million.’

The frequency of occurrence of the indefinite articles *bir* and *bitta* in the novel from 2001 correlates with the distinction between literary and colloquial language: *bir* is significantly more frequent in the author's narrative, whereas *bitta* is significantly more frequent in direct speech.

We can therefore conclude that at least in the written register, there is an obvious development of the use of *bitta* as an indefinite article. The fact that our corpus was limited in size as well as in kind precludes any claims about the use of *bitta* as an indefinite article in spoken Uzbek around 1926. We simply cannot decide, based on our corpus results, whether *bitta* was used neither in written nor spoken Uzbek around 1926 or whether *bitta* was used, but only in spoken Uzbek. What is plausible, however, is that *bitta* has developed into an indefinite article after *bir*. The first indication that this might be the case is that the written register usually lags behind the spoken register, and if *bitta* had developed before *bir* we would have expected at least as many occurrences of *bitta* as of *bir* in the novel from 1926. The second indicator that *bitta* has developed after *bir* is that we found 9 occurrences of *bir* in a predicative construction in the novel from 2001, while *bitta* has not occurred in predicative constructions in this novel. To the extent that we have reason to believe that the item which develops first into an indefinite article also reaches the last stage of development first, we also

have reason to believe that *bir* developed first into an indefinite article, since it appears to have reached the last stage of development first. Moreover, due to the fact that in the novel from 2001 three quarters of the occurrences of *bir* are in the author's narrative, whereas nine tenths of the occurrences of *bitta* are in the direct speech of the characters, it is also plausible to hypothesise a significant difference between written and spoken Uzbek: whereas in written Uzbek *bir* is more widely used as an indefinite article, in spoken Uzbek it is *bitta* which is more widely used as an indefinite article.

Putting together the hypothesis that *bir* has started developing into an indefinite article before *bitta* with the hypothesis that in modern spoken Uzbek *bitta* is more widespread than *bir*, leads to the conclusion that the use of *bitta* as an indefinite article has spread in spoken Uzbek to the expense of *bir*. We have still to add the caveat of the very limited corpus and of the decrease in the overall usage of these articles: while in the 1926 novel 10% of the indefinite NPs contained the indefinite article *bir*, in the 2001 novel only 2,7% of the indefinite NPs contained *bir* or *bitta*.

In the next section we provide additional evidence from an online questionnaire bearing on this issue. First we wanted to know whether, and if so in which contexts of use, in modern spoken Uzbek the use of *bitta* is judged better than the use *bir*. Secondly, since the absence of evidence from our corpus that *bitta* has reached stage 5 in the development from numeral to indefinite article is not evidence that it has not reached this stage, the question arises how good sentences are judged in which the indefinite article in a predicative construction is not *bir* but *bitta*.

5. Questionnaire

To answer these questions we designed and carried out a questionnaire eliciting acceptability judgements about sentences containing the two articles in the four different usages. The independent factors of our experiment were first the indefinite article, namely *bir* or *bitta*, secondly the animacy of the referent (human or inanimate), and thirdly the four different usages (stage 2 to 5). Crossing the values of these three factors with each other, we got 16 different types of sentences which were instantiated in the questionnaire by means of 3 different lexicalisations. The total of 48 items was divided onto 6 actual questionnaires, so that each participant judged only 8 items which were randomly mixed with 20 filler

sentences. The task of the 45 participants in the questionnaire was to judge the acceptability of the sentences (“how good the sentence sounds”) on a scale from 2, meaning very bad, to 5, meaning very good.

The judgements were collected in part online (using the WEBEXP2 software developed in Edinburgh) and in part by means of written questionnaires, and were then evaluated by means of an analysis of variance (ANOVA), as illustrated in Figure 1.⁹

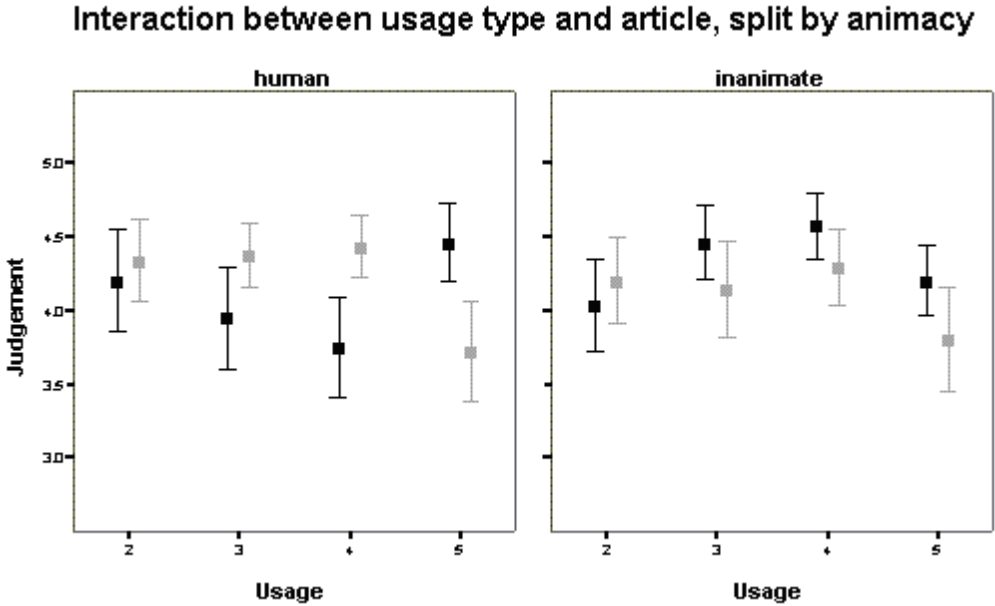


Figure 1: Interaction between usage and article, split by animacy.
(Black error bars = *bir*, grey error bars = *bitta*)

First, we found a significant interaction (with $F(3,561) = 4.7$; $p < 0.005$) between the three factors usage, article and animacy, as the two charts in table 4 show. The error bars indicate that if we repeated the experiment with other participants the likelihood of the mean of the judgements being between the upper and lower bounds of the error bars is 95%. Therefore, if two error bars do not overlap, then the likelihood is very high that the difference is not due to chance but statistically significant. The black error bars indicate the mean judgements for *bir*,

⁹ The basic idea behind an ANOVA is to estimate how likely or unlikely it is for a given difference in judgements to be due to chance. If it is unlikely that the difference is due to chance, then the difference is called statistically significant. This estimation is based on comparing the variance between two (or more) sets of judgements (e.g. the set of judgements of sentences whose indefinite article is *bir* with the set of judgements whose indefinite article is *bitta*) with the variance within these sets of judgements. If the variance between the sets of judgements is high while the variance within the sets of judgements is low, then the difference is likely to be statistically significant. If on the other hand, the variance between the sets of judgements is low, while the variance within the sets of judgements is high then the difference is likely to be due to chance.

whereas the grey error bars indicate the mean judgements for *bitta*. The horizontal axis represents the four different types of usages (2 = introduction of referent which is picked up again, 3 = introduction of referent which is not picked up again, but that is specific, 4 = referential identity unknown by speaker, 5 = predicative construction), and the vertical axis represents the judgements from 2 (very bad) to 5 (very good).

We can make four main observations concerning the usage of *bitta*, the influence of animacy, the distribution along Heine's scale and the contrast between argumental vs. non-argumental indefinites.

1. Except for the use of *bitta* in predicative constructions, all other uses of *bitta* as an indefinite article have a mean judgement of over 4 on a scale from 2 to 5, irrespective of the animacy of the referent. The judgement mean for sentences containing *bitta* in predicative constructions is around 3.75, and is thus clearly above the mean for ungrammatical filler sentences, which was around 2.5. So the hypothesis from the previous section, namely that *bitta* is more widespread in spoken Uzbek than it is in written Uzbek, is supported by the fact that with the exception of stage 5 uses it is also consistently judged to be as acceptable as the grammatical filler sentences. Moreover, the fact that the stage 5 use of *bitta* is clearly more acceptable than the ungrammatical sentences, but not as good as *bitta* in its other uses as an indefinite article indicates that the use of *bitta* is in the process of spreading to the fifth stage of development into an indefinite article.

2. There is a clear influence of animacy on the choice of the indefinite article: *bir* is more acceptable with inanimate nouns and *bitta* is preferred with animate nouns (except for stage 5) – this result corresponds to the corpus analysis of the 2001 novel, where *bir* was used 4,4 times more often with inanimate than with animate nouns, and *bitta* was used only 2,4 times more often with inanimate nouns than with animate ones. However, the result contradicts some of the judgements of our informant.

3. Analysing the distribution of *bir* and *bitta* along the evolutionary scale (2: pragmatic discourse salient, topical; 3: specific, but not necessarily topical, 4: introduces discourse referent, but not specific), we see for animate nouns that *bitta* is clearly better than *bir*, but that the difference between stage 3 and 4 is not significant. This correlates with the results from the corpus search of the 2001 novel, where the distribution for *bir* and *bitta* with animate nouns for stages 2-4 is nearly identical. We see for inanimate nouns, that the use of *bitta* is slightly preferred. If we summarize these observations as in Table 6:

Table 6 Comparing the use of *bir* and *bitta* on the evolutionary scale and animacy

state	2 (topical)	3 (specific)	4 (discourse referent)	5 predicational
animate	bir = bitta	bitta (< bir)	bitta (< bir)	bir (*bitta)
inanimate	bir = bitta	bir < bitta	bir < bitta	bir < bitta

The table clearly indicates that there is no contrast between specific (in the sense of “speaker known”) and non-specific indefinites as assumed in section 3.4 above, but only a contrast between usages where the indefinite introduces a discourse referent (stages 2-4) and ones where it does not (stage 5: predicational or non-argumental use). A second observation concerns the evolution of the two indefinite articles: At a first glance, this table seems to differ from grammaticalization processes depending on “individuation”, i.e. animacy and the referentiality scale. However, on closer inspection it is not so clear whether there is such an evolutionary development or just a reflection of particular properties of *bir* and *bitta*. This connects with the next observation.

5. The strongest contrast we found is that *bitta* cannot be used for animate nouns in predicational positions – this is corroborated by the corpus analysis, where we did not find any instance of *bitta* in predicational position. This may provide the motivation for the semantic contrast observed in examples (27) and (28) in section 3.4, where the use of *bir* implies a non-specific reading, whereas the use of *bitta* is unmarked. Thus the contrast corresponds to a different concept of specificity, namely referential specificity that is only defined in opaque contexts: the specific reading introduces an argument, while the non-specific reading introduces a property into the sentence. One interpretation of the observation is that *bitta* is just not developed as far as *bir*, another interpretation is that *bitta* has a function that does not allow predicative interpretation.

6. Conclusion

In this paper we have provided converging evidence from (i) our informant, (ii) from a corpus study and (iii) from an online questionnaire that modern Uzbek has two indefinite articles *bir* and *bitta*. The corpus study revealed a significant difference in usage between *bir* and *bitta*, to the effect that *bir* was strongly preferred in the author’s narrative, whereas *bitta* was preferred

in direct speech. On the basis of this we hypothesised that *bitta* has spread in modern colloquial Uzbek to the expense of *bir*. This hypothesis was partly confirmed by the evaluation of an online questionnaire, which indicates that in certain contexts of use *bitta* is judged more acceptable than *bir*. Moreover, the acceptability judgements on sentences containing *bitta* in predicative position suggests that the use of *bitta* is in the process of spreading to the last stage of development into an indefinite article. There are many open questions with respect to the conclusions but also with respect to the application of the evolutionary scale to the Uzbek data: The percentage of indefinite noun phrases with *bir* or *bitta* decreased from about 10% in the 1926 novel to about 4% in the 2001 novel. This indicates that the articles are far from becoming obligatory, they are still optional. In other words, we do not have a system with two oppositions, but with three – unmarked (or bare), *bir* and *bitta* (very similar to the facts in Moroccan, see section 2.1). The distribution of the two indefinite articles is quite balanced (except for the predicational stage 5), which also contradicts a goal-oriented development. The two main contrasts of the corpus studies, namely the dependency on register (narrative vs. oral presentation) and the contrast in animacy need further explanation and do not fit into a classical picture of indefiniteness division as suggested in different scales. The diachronic perspective has helped us arrive at a richer and broader picture, but due to the small size of the corpus we cannot form very strong hypotheses. The synchronic study with the help of questionnaires provided additional data, which again are not easy to interpret. We need a more detailed study of the semantics of each of the indefinite articles, in particular of *bitta*, as the literature suggests that the suffix *-ta* might be used to discourse-link the number it introduces, which would explain the ungrammaticality in predicative constructions. What we really have learned from this fascinating distribution of the two indefinite articles in Uzbek is that the category of indefiniteness is quite complex and that markers for indefiniteness might establish contrasts that do not necessarily follow from our semantic reconstruction of indefiniteness.

Abbreviations

ABL – ablative	F – feminine	PROG – progressive
ACC – accusative	FUT – future	PRS – present
AGR – agreement	GEN – genitive	PTCP – participle
CLF – classifier	IMP – imperative	SG – singular
CVB – converg	LOC – locative	SPEC – specifier (in contrast to a classifier)
COND – conditional	NEG – negation	SPRL - superlative
DAT – dative	PL – plural	
DEM – demonstrative	POSS – possessive	
EVID – evidential	PRF – perfect	

References

- Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Bauer, W. (1993). *Maori*. London and New York: Routledge.
- Beckwith, C. (1998). Noun Specification and Classification in Uzbek. *Anthropological Linguistics* 40, 124–140.
- Bodrogligeti, A. J. E. (2003). *An Academic Reference Grammar of Modern Literary Uzbek*. 2 Volumes. München: Lincom Europa.
- Boeschoten, H. (1998). Uzbek. In L. Johanson, and É. Á. Csató (Eds.), *The Turkic Languages*, pp 357–378. London: Routledge.
- Brustad (2000) auf Seite 4 gefunden
- Dryer, M. S. I. (2011). Indefinite Articles. Chapter 38. The World Atlas of Language Structures Online. Available at <http://wals.info/chapter/1>. Accessed 23 September 2011.
- Ebert, C. and S. Hinterwimmer (2012) this volume
- Farkas, D. (2002). Specificity Distinction. *Journal of Semantics* 19, 213–243.
- Fassi-Fehri, A. (2006). *A short note on Moroccan specifics*. Ms., University of Rabat.
- von Gabain, A. (1945). *Özbekische Grammatik*. Mit Bibliographie, Lesestücken und Wörterverzeichnis. Leipzig und Wien: Harrassowitz.
- Givón, T. (1973). Opacity and reference in language: an inquiry into the role of modalities. In J. Kimball (Ed.), *Syntax and Semantics 2*, pp. 95–122. New York: Academic.
- Givón, T. (1981). On the Development of the Numeral 'one' as an Indefinite Marker. *Folia Linguistica Historica*

2, 35–53.

- Harrell, R. S. (1962). *A short reference grammar of Moroccan Arabic*. Washington D.C.: Georgetown University Press.
- Haspelmath, M. (1997). *Indefinite pronouns*. (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.
- Heine, B. (1997). *Cognitive Foundations of Grammar*. Oxford: Oxford University Press.
- von Heusinger, K. (2011a). Specificity. In K. von Heusinger, C. Maienborn, and P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning. Volume 2*, pp. 1025–1058. Berlin: de Gruyter.
- von Heusinger, K. (2011b). Specificity, referentiality and discourse prominence: German indefinite demonstratives. In I. Reich et al. (Eds.), *Proceedings of Sinn & Bedeutung 15*, pp. 1–22. Saarbrücken: Universaar – Saarland University Press.
- Keenan, E., and K. Ebert (1973). A Note in Marking Transparency and Opacity. *Linguistic Inquiry* 4, 421–424.
- Kornfilt, J. (1997). *Turkish*. Descriptive Grammars. London: Routledge.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Malik, T. (2001). *Shaytanat*. Part 4. Ort: Verlag???
- Niyazmetova, D. (2009). *Differential object marking in Uzbek*. Master thesis, University of Stuttgart.
- Qodiriy, A. (1926). *Otgan kunlar*. Ort: Verlag??
- Schroeder, C. (2006). Articles and Article Systems in some Areas of Europe. In G. Bernini, and M. Schwartz (Eds.), *Pragmatic Organization of discourse in the language of Europe*, pp. 545–611. Berlin: Mouton de Gruyter.
- Schwarz, F. (2009). Two Types of Definites in Natural Language. Dissertation, University of Massachusetts, Amherst.
- Wright, S., and T. Givón (1987). The pragmatics of indefinite reference: Quantified text-based studies. *Studies in Language* 11, 1–33.
- Yanovich (2012) this volume