Nominal anchoring

Specificity, definiteness and article systems across languages
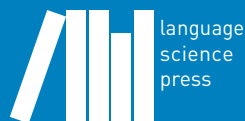
Edited by

Kata Balogh

Anja Latrouite

Robert D. Van Valin, Jr.

Topics at the Grammar-Discourse Interface ??

language science press

Topics at the Grammar-Discourse Interface

Editors: Philippa Cook (University of Göttingen), Anke Holler (University of Göttingen), Cathrine Fabricius-Hansen (University of Oslo)

In this series:

1. Song, Sanghoun. Modeling information structure in a cross-linguistic perspective.

2. Müller, Sonja. Distribution und Interpretation von Modalpartikel-Kombinationen.

3. Bueno Holle, Juan José. Information structure in Isthmus Zapotec narrative and conversation.

4. Parikh, Prashant. Communication and content.

Nominal anchoring

Specificity, definiteness and article systems across languages

Edited by

Kata Balogh

Anja Latrouite

Robert D. Van Valin, Jr.

Contents

Chapter 1

Nominal anchoring: Introduction

Kata Balogh
Heinrich-Heine-Universität Düsseldorf

Anja Latrouite
Heinrich-Heine-Universität Düsseldorf

Robert D. Van Valin, Jr.
Heinrich-Heine-Universität Düsseldorf & University at Buffalo

1 Aims and motivations

It has been observed that a multitude of the world's languages can do without explicit formal marking of the concepts of definiteness and specificity through articles (e. g., Russian, Tagalog, Japanese), while other languages (e. g., Lakhota) have very elaborate systems with more fine-grained distinctions in the domains of definiteness and specificity-marking. The main questions that motivate this volume are: (1) How do languages with and without an article system go about helping the hearer to recognize whether a given noun phrase should be interpreted as definite, specific or non-specific? (2) Is there clear-cut semantic definiteness without articles or do we find systematic ambiguity regarding the interpretation of bare noun phrases? (3) If there is ambiguity, can we still posit one reading as the default? (4) What exactly do articles in languages encode that are not analyzed as straightforwardly coding (in)definiteness? (5) Do we find linguistic tools in these languages that are similar to those found in languages without articles?

Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr.

The papers in this volume address these main questions from the point of view of typologically diverse languages. Indo-European is well represented by Russian, Persian, Danish and Swedish, with diachronic phenomena investigated in relation to the last two of these. In terms of article systems, they range from Russian, which has no articles, the typical situation in most Slavic languages, to Persian, which has an indefinite article but no definite article, to the more complete systems found in Romance and Germanic languages. The three non-Indo-European languages investigated in this volume, namely Mopan (Mayan) , Vietnamese and Siwi (Berber), are typologically quite diverse: Mopan is verb-initial and thoroughly head-marking, Vietnamese is verb-medial and radically isolating, i. e., lacking inflectional and derivational morphology, and Siwi is verb-initial with the signature Afro-Asiatic trilateral roots which are the input to derivational and inflectional processes. What they have in common is the absence of articles signaling (in)definiteness.

2 Article systems and related notions

Chesterman (1991) points out that "it is via the articles that definiteness is quintessentially realized, and it is in analyses of the articles that the descriptive problems are most clearly manifested. Moreover, it is largely on the basis of the evidence of articles in article-languages that definiteness has been proposed at all as a category in other languages."

Here, we view definiteness as a denotational, discourse-cognitive category, roughly understood as identifiability of the referent to the speaker, instead of a grammatical (or grammaticalized) category, and therefore we can investigate the means that languages use for indicating definiteness or referential anchoring in general. Natural languages have various means to signal definiteness and/or specificity. Languages differ in their article systems as well as in the functions the set of articles they exhibit may serve. Simple article languages (e. g., English, Hungarian) generally distinguish definite and indefinite noun phrases by different articles, but they may also use their article inventory to code categories other than definiteness (e. g., Mopan Maya). Complex article languages like Lakhota, which exhibits an elaborate and sophisticated system, always mark more than simple (in)definiteness. A great number of languages (e. g., Russian, Tagalog, Japanese) have no or no clear-cut article systems and rely on other means to encode definiteness distinctions.

Most of the languages investigated in this volume belong to the last type. The means they use to help indicate how the referent of a noun phrase is anchored

and intended to be interpreted include classifier systems (e. g., Vietnamese, Chuj), clitics (e. g., Romanian), designated morphemes on nouns (e. g., Moksha, Persian) and syntactic position (e. g., Chinese). In certain languages, alongside article systems and morphosyntactic means, prosody plays a crucial role for the coding of (in)definiteness, for example accent placement in Siwi or tone in Bambara.

2.1 Basic notions: definiteness and specificity

In the cross-linguistic investigation and analysis of article systems and noun phrases, various different but related notions play a key role. In the analysis of various types of definite and indefinite noun phrases, the two most important notions are *definiteness* and *specificity*, together with further distinguishing notions of *uniqueness*, *familiarity*, *discourse prominence* and so on. In the following we give a brief introduction to these notions. Our aim is not to provide a detailed discussion of all notions and all theories, but to present an overview of the most important classical analyses relevant to the papers and their main issues in this volume.

2.1.1 Definiteness

The notion of *definiteness* itself is a matter of controversy, given the different uses of definite noun phrases for anaphoric linkage, relational dependencies, situational/deictic salience or inherently uniquely referring nouns. The notion is used in a variety of ways by different authors. The classical analyses of definiteness distinguish two main lines of characterization: (1) the *uniqueness* analysis, following works by Russell (1905) and Strawson (1950), and (2) the *familiarity* account, after Christophersen (1939), Kamp (1981) and Heim (1982).

In Russell's (1905) analysis, indefinites have existential quantificational force, while definite descriptions[1] are considered referential. Definites assert *existence* and *uniqueness*, as illustrated in the logical translation of sentences like (1).

(1) The N is P.
 $\exists x(N(x) \land \forall y(N(y) \to x = y) \land P(x))$
 a. there is an N (existence)
 b. at most one thing is N (uniqueness)
 c. something that is N is P

The meaning contribution of the definite article is to signal the existence of a unique referent (a-b), while the head noun provides sortal information of

---

[1]These mostly refer to noun phrases with a definite article, e. g., *the dog*, but other expressions like possessive noun phrases and pronouns are also considered definite descriptions.

the referent (c). In the Russellian tradition, indefinites are distinguished from definites in terms of uniqueness, as the predicate (sortal information) applies to exactly one referent. Russell's highly influential approach has inspired many theories on definiteness; similarly, various approaches point out critical issues in Russell's theory. The most intriguing issues discussed in the literature are the problem of presuppositionality, the problem of incomplete descriptions and the problem of referentiality. To solve these crucial issues a great number of theories have been proposed over the decades. In Strawson's (1950) account, existence and uniqueness are presupposed rather than asserted. He claims that if the presupposition fails, the sentence does not bear a truthvalue, i. e., it is neither true nor false. The incompleteness problem, where the definite description does not have a unique referent, inspired several authors (e. g., Strawson 1950; McCawley 1979; Lewis 1979; Neale 1990) to offer various solutions, like contextual restriction and the prominence/saliency approach. The latter was proposed by McCawley (1979) and Lewis (1979), who argues that definite descriptions refer to the most prominent or most salient referent of a given context. Donnellan (1966) argues that definite descriptions have two different uses: an attributive and a referential use. The former use can be characterized similarly to Russell's account, while the latter use requires a different analysis. Donnellan's famous example is (2), which can be used in different ways in different situations.

(2)   Smith's murderer is insane.   (Donnellan 1966: p.285)

In a situation where the murderer is unknown (e. g., at the scene of the crime), the noun phrase 'Smith's murderer' is understood attributively as meaning that whoever murdered Smith is insane. On the other hand, in a different situation where the murderer is known (e. g., at the trial), the noun phrase can be replaced by, for example, he, as it is used referentially, referring to the individual who is the murderer.

The other highly influential classical account of definites represents a different view. These theories follow the work by Christophersen (1939), who accounts for the interpretation of definites in terms of *familiarity* rather than uniqueness. In his theory, definite descriptions must be discourse-old, already introduced in the given discourse context, and as such known to the hearer. Christophersen's familiarity account inspired famous theories in formal semantics: File Change Semantics [FCS] of Heim (1982) and the similar Discourse Representation Theory [DRT], which was developed in parallel and introduced by Kamp (1981) and Kamp & Reyle (1993). One of the major contributions of these approaches is the solution for the so-called 'donkey sentences' (3a), and further issues of the interpretation

of discourse anaphora (3b).

(3) a. If a farmer$_i$ owns a donkey$_j$, he$_i$ beats it$_j$.
     b. A student$_i$ came in. She$_i$ smiled.

In both sentences, the indefinite noun phrases can be referred to by an anaphoric expression in the subsequent sentence. Based on such examples, they propose a division of labour between indefinite and definite noun phrases. Indefinites like *a student* introduce new discourse referents, while definite noun phrases like *the student* pick up a referent that has already been introduced, similarly to anaphoric pronouns.

In his 1985 paper, Löbner argues for a relational approach and against the uniqueness approach, claiming that it is rather non-ambiguity that is essential for definiteness. Löbner (1985) distinguishes noun phrases by their type of use. The main distinction is into *sortal* and *non-sortal* nouns, where the latter is further divided into *relational* and *functional* nouns and concepts. Relational nouns include kinship terms (e. g., *sister*), social relations (e. g., *friend*) and parts (e. g., *eye*), while functional nouns are roles (e. g., *wife, president*), unique parts (e. g., *head, roof*), conceptual dimensions (e. g., *height, age*) and singleton events (e. g., *birth, end*). In the analysis of definite descriptions, Löbner (1985) distinguishes semantic and pragmatic definites. For semantic definites "the referent of the definite is established independently of the immediate situation or context of the utterance" (Löbner 1985: p.298), while pragmatic definites are "dependent on special situations and context for the non-ambiguity of a referent" (Löbner 1985: p.298). One- and two place functional concepts (4), as well as configurational uses (5), are considered semantic definites. Löbner claims that statements like (5) are impossible with sortal nouns.

(4) 1-place functional concepts: *the time, the last party, the other girl*, etc.
     2-place functional concepts: *my wife, the author, the president of France*, etc.

(5) He was the son of a poor farmer.   (Löbner 1985: ex.17)

As Löbner argues, this distinction is significant in various ways; for example, functional nouns can only take the definite article (with the exception of existential contexts). Further examples he gives are of German cliticization (6), where the cliticized article encodes a semantic definite as opposed to a non-cliticized one. In various languages, there are different articles, often referred to as weak and strong (Schwarz 2019), encoding semantic and pragmatic definites.

Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr.

This distinction can be found, for example, in the Fering (Föhr) dialect of Frisian (e. g., Ebert 1971) and in the Rheinland dialect of German (e. g., Hartmann 1982).

(6) **German**:
  a. Er muß ins/*in das   Krankenhaus.
     he must in=the/*in the hospital
     'He must go to hospital again.'
                              (from Löbner 1985: ex.52, glosses are from us)
  b. Er muß wieder *ins/in das   Krankenhaus zurück, aus dem   er
     he must again   in=the/in the hospital   back from the.DAT he
     schon   entlassen war.
     already discharged was
     'He has to go back to the hospital from which he already been
     discharged.'   (from Löbner 1985: ex.54, glosses are from us)

(7) **Fering (Föhr)**:   (Ebert 1971: p.161)
  a. Ik skal deel tu a/*di   kuupman.
     I   must down to the_W/the_S grocer
     'I have to go down to the grocer.'
  b. Oki hee an hingst keeft.   *A/di   hingst haaltet.
     Oki has a horse bought the_W/the_S horse limps
     'Oki has bought a horse. The horse limps.'

As for the meaning contribution of the definite article, based on the different noun/concept types and their uses, Löbner (1985) argues that the definite article indicates that the given noun must be taken as a functional concept.

2.1.2 Specificity

The notion of *specificity* (see, e. g., von Heusinger 2011) is also defined and characterized in different ways and in relation to a variety of factors. Specificity is generally used to distinguish various readings of indefinites. A generally accepted view is that sentences like (8) can be interpreted in two ways, depending on whether the speaker has a particular entity in mind, referred to by the indefinite noun phrase.

(8) *Mia kissed a student yesterday.*
    1. whoever Mia kissed is a student (non-specific)
    2. there is a specific student whom Mia kissed (specific)

As a linguistic notion, the opposition between the non-specific and the specific readings of indefinites is characterized in relation to a variety of factors. Farkas (1994) distinguishes referential, scopal and epistemic specificity. Specific indefinites refer to an individual, and hence can be anaphorically referred back to. With respect to the second reading of (8), the sentence could be followed by *He is tall*, while this is not possible after the first reading. In relation to other operators, specific indefinites take a wide scope. The epistemic opposition is very close to (if not the same as) the referential opposition, as it is characterized by the fact that, by using specific indefinites, the speaker has a referential intention, i. e., they have a certain individual in mind (Karttunen 1968; Farkas 1994). In addition to Farkas's (1994) three-way distinction, von Heusinger (2011; 2019) proposes four more oppositions, namely partitivity, noteworthiness, topicality and discourse prominence. As Enç (1991) argues, specific indefinites are discourse-linked and inferable: they refer to a part of a set previously introduced to the discourse. As a motivation, she shows that this distinction is overtly marked in Turkish: accusative marked direct objects are interpreted specifically (9a), while unmarked objects are taken as non-specific (9b).

(9) a. Ali bir piyano-yu kiralamak istiyor.
Ali one piano-ACC to-rent   wants
'Ali wants to rent a certain piano.'   (Enç 1991: ex.12)

b. Ali bir piyano kiralamak istiyor.
Ali one piano   to-rent   wants
'Ali wants to rent a (non-specific) piano.'   (Enç 1991: ex.13)

The relevance of noteworthiness is often illustrated by the use of the marked indefinite *this N* construction. Such examples can only be followed by newsworthy/interesting/particular information regarding the noun phrase.

(10) He put this 31-cent stamp on the envelope,   (after Maclaran 1982)
a. and only realized later that it was worth a fortune.
b. #so he must want it to go airmail.

Topicality and discourse prominence are also closely related to specificity. Indefinite noun phrases that are topical receive a specific interpretation. This can be shown by Hungarian examples, where topicality is syntactically marked by placement to a left-peripheral position within the clause (11).

Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr.

(11) Egy diák   be-kopogott   az igazgató-hoz.
     a   student vprt-knocked the director-ALL
     'A (particular) student knocked at the director's office.'

The left-peripheral topic position can only host referential and specific noun phrases (e. g., É. Kiss 2002), and hence the indefinite noun phrase can only be in the topic position when it is interpreted specifically.

3 Contributions

The papers in this volume address to different degrees the general questions introduced in section 1. Most contributions report on research on different corpora and elicited data or present the outcome of various experimental studies. One paper presents a diachronic study of the emergence of article systems. As mentioned before, the volume covers typologically diverse languages: Vietnamese, Siwi (Berber), Russian, Mopan (Mayan), Persian, Danish and Swedish.

3.1 Languages with articles

If a language is analyzed as having an article, the standard expectation is that it will express either definiteness or indefiniteness. However, the number of papers introducing article-languages in which the determiners do not encode different degrees of identifiability and uniqueness is on the rise (e. g., Lyon 2015). The crucial question is what features an element is required to exhibit to be counted as an article. If the answer is given in line with Himmelmann (2001) and others, then no functional element that does not convey some degree of specificity is counted as an article. If Dryer's (2014) characterization of articles is adopted, then all functional elements that occur with high frequency in noun phrases, indicate argumenthood and vary for grammatical features are included in the category.

Eve Danziger and Ellen Contini-Morava adopt Dryer's (2014) view in their contribution *Referential anchoring without a definite article: The case of Mopan (Mayan)* and investigate all the means that Mopan utilizes in order to evoke relative identifiability and uniqueness. While, based on formal and distributional criteria, the Yucatecan language Mopan exhibits a determiner of the type usually classified as an article, they find that this article does not encode any of the semantic notions of definiteness, specificity and uniqueness. It merely serves to express that a given lexeme is used as an argument in the sentence. In their analysis and explanation, they build upon Dryer's (2014) definiteness hierarchy and demonstrate that the article itself, as well as the bare nominal form, can occur

in any position in Dryer's definiteness hierarchy. This observation leads to an investigation of exactly what the discourse-pragmatic function of the article is and how it can be calculated. The authors' conclusion is that the contribution of the article is best characterized by factors such as discourse salience, which contexts or world knowledge may lend even to non-specific indefinites.

In their paper, *The specificity marker -e with indefinite noun phrases in Modern Colloquial Persian*, Klaus von Heusinger and Roya Sadeghpoor focus on the specificity marker *-e* and its compatibility with two indefinite markers and investigate the different kinds of indefinite readings that arise. In their experimental pilot studies, they test and provide some support for the hypothesis that the difference in interpretation between the combinations lies in the anchoring of the referents, i. e., in whether the referent is construable as speaker-specific or non-speaker-specific. The studies thereby provide additional evidence for the need to assume a fine-grained approach in the investigation of specificity and referential anchoring (von Heusinger 2002). However, they also show that specificity-unrelated semantic properties like *animacy* need to be taken into account in the explanation of their results.

The contribution *Indirect anaphora from a diachronic perspective: The case of Danish and Swedish* by Dominika Skrzypek is the only diachronic study in this volume. The author investigates different kinds of indirect anaphora (*associative anaphora, bridging anaphora*) as one of the steps in the grammaticalization process towards a definite article from the beginning of the 13th century until the middle of the 16th century. The paper is particularly concerned with the distribution and use of indirect anaphora and the features that the relationship between indirect anaphora and their anchor is based on. Looking at inalienable and other types of indirect anaphora, the author shows that indirect anaphora form a heterogeneous concept and are not easily positioned in the strong-weak definiteness dichotomy. The evidence points to the fact that the definite article did not spread uniformly through indirect anaphora in Danish and Swedish.

3.2 Languages without articles

In article-less languages, the encoding of definiteness is often a complex matter, where various linguistic factors play a role. Japanese and Chinese are both languages that are well known for lacking an article system. In Japanese, argument phrases are marked by case markers (nominative: *ga*, accusative: *wo*, dative: *ni*) or non-case markers like the topic marker *wa* or the additive marker *mo* 'also'. Consequently, definiteness is not straightforwardly grammaticalized, but rather considered an interpretational category (e. g., Tawa 1993), for which classifiers

Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr.

play a crucial role. The same holds for Chinese, which lacks case markers, but exhibits even more numeral classifiers than Japanese. These have been argued by Cheng & Sybesma (1999) and others to play a crucial role for the definiteness reading of noun phrases, whenever numeral information is missing. However, Peng (2004: p.1129) notes that for indeterminate expressions "there is a strong but seldom absolute correlation between the interpretation of identifiability or nonidentifiability and their occurrence in different positions in a sentence". Simpson et al. (2011), who study bare classifier definites in Vietnamese, Hmong and Bangla, also find that classifiers are relevant to nominal anchoring. However, the fact that bare noun phrases also seem to be able to receive definite interpretations weakens the claim that classifiers are the morphosyntactic key to definiteness interpretation, and rather points to the fact that a multilevel approach proposed by Heine (1998) is better in explaining how definiteness or specificity interpretations arise.

Walter Bisang and Kim Ngoc Quang, in their study *(In)definiteness and Vietnamese classifiers*, contribute to our understanding of the classifier language Vietnamese. They investigate which linguistic factors influence the interpretation of phrases with numeral classifiers [CL] in bare classifier constructions as either definite or indefinite and point out the licensing contexts for the different uses and readings of nominal classifiers. They find a striking clustering of definite interpretations with animacy and subject status, whereby definiteness is understood as identifiability in discourse. Indefinite interpretations, on the other hand, are predominantly witnessed in certain sentence types (existential sentences and thetic sentences) and with certain types of verbs (verbs of appearance). A crucial finding is that noun class type, following Löbner (1985; 2011), and factors like *animacy* and *grammatical relation* are less important than information structure for the appearance of classifiers in definite and indefinite contexts. Classifiers are shown to be associated with pragmatic definiteness, rather than semantic definiteness, i. e., identifiability rather than uniqueness. Furthermore, the authors provide evidence that contrastive topics, contrastive focus and focus particles correlate with the use of classifier constructions. Similar to the constructions discussed for Persian and Mopan (Mayan) in the volume, the classifier construction in Vietnamese can be once more viewed as a construction whose final interpretation depends, on the one hand, on discourse prominence and, on the other hand, on features of the morphosyntax-semantics interface that are well known for contributing to the overall saliency of a phrase.

In her contribution, *Accent on nouns and its reference coding in Siwi Berber*, Valentina Schiattarella investigates definiteness marking in Siwi Berber, an indigenous Berber language spoken in Egypt. In Siwi, a language without articles,

it is claimed that the placement of accent on the last syllable versus the penultimate syllable encodes indefiniteness and definiteness respectively, i. e., the accent on the last syllable is generally assumed to encode indefiniteness and the accent on the penultimate syllable to encode definiteness. This default interpretation can be overridden, as Schiattarella shows in her paper. She analyzes various corpus data from spontaneous discourse and guided elicitations to further examine the role of various morphosyntactic means (e. g., possessive constructions, demonstratives, prepositions and adpositional phrases) as well as pragmatic aspects (e. g., anaphoricity, familiarity, uniqueness, reactivation and information structural considerations) in influencing the interpretation of noun phrases. The author, furthermore, finds that right- and left-detached constructions or the appearance of a demonstrative, a possessive marker or relative clause in postnominal position influences the interpretation.

Olga Borik, Joan Borràs-Comes and Daria Seres, in *Preverbal (in)definites in Russian: An experimental study*, present an experimental study on Russian bare nominal subjects, and investigate the relationship between definiteness, linear order and discourse linking. Given that Russian lacks articles and has very flexible word order, it is widely assumed that (in)definiteness correlates with the position of a noun phrase in the clause, i. e., preverbal position is associated with a definite reading and postverbal with an indefinite interpretation. The authors experimentally verify that this correlation basically holds, but they also find that speakers accept a surprising number of cases in which a preverbal NP is interpreted as indefinite, which leads to the conclusion that Russian bare nouns are basically indefinite. The unexpected correlations between position and interpretation lead to further investigations of the relevant factors involved and the suggestion that, regardless of topicality, discourse linking principles following Pesetsky (1987) and Dyakonova (2009) facilitate the use of indefinite nominals in the unexpected preverbal position.

3.3 Summary

The papers in this volume deal with pragmatic notions of definiteness and specificity. The studies presented here provide the following findings regarding our initial motivating questions. On the issue of how languages with and without articles guide the hearer to the conclusion that a given noun phrase should be interpreted as definite, specific or non-specific, the studies in this paper argue for similar strategies. The languages investigated in this volume use constructions and linguistic tools that receive a final interpretation based on discourse prominence considerations and various aspects of the syntax-semantics interface. In

Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr.

case of ambiguity between these readings, the default interpretation is given by factors (e. g., familiarity, uniqueness) that are known to contribute to the salience of phrases, but may be overridden by discourse prominence. Articles that do not straightforwardly mark (in)definiteness encode different kinds of specificity. In the languages studied in this volume, whether they have articles or do not have an article system, we find similar factors and linguistic tools in the calculation process of interpretations.

Acknowledgments

References

Cheng, Lisa & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4). 509–542.

Chesterman, Andrew. 1991. *On definiteness: A study with special reference to English and Finnish* (Cambridge Studies in Linguistics 56). Cambridge: Cambridge University Press.

Christophersen, Paul. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Einar Munksgaard.

Donnellan, Keith S. 1966. Reference and definite descriptions. *The Philosophical Review* 75(3). 281–304.

Dryer, Matthew S. 2014. Competing methods for uncovering linguistic diversity: The case of definite and indefinite articles (commentary on Davis, Gillon, and Matthewson). *Language* 90(4). e232–e249.

Dyakonova, Marina. 2009. *A phase-based approach to Russian free word order*. Amsterdam: University of Amsterdam. (Doctoral dissertation).

É. Kiss, Katalin. 2002. *The syntax of Hungarian.* Cambridge: Cambridge University Press.

Ebert, Karen. 1971. *Referenz, Sprechsituation und die bestimmten Artikel in einem nordfriesischen Dialekt* (Studien und Materialien Nr. 4.). Bredstedt: Nordfriisk Instituut.

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

Farkas, Donka. 1994. Specificity and scope. In Lea Nash & George Tsoulas (eds.), *Proceedings of langues et grammaire 1*, 119–137.

Hartmann, Dietrich. 1982. Deixis and anaphora in German dialects: The semantics and pragmatics of two definite articles in dialectal varieties. In Jürgen Weissenborn & Wolfgang Klein (eds.), *Here and there: Cross-linguistic studies on deixis and demonstration* (Pragmatics & Beyond III:2-3), 187–208. Amsterdam/Philadelphia: John Benjamins.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases.* Arnhem: University of Massachusetts. (Doctoral dissertation).

Heine, Julia E. 1998. Definiteness predictions for Japanese noun phrases. In Christian Boitet & Pete Whitelock (eds.), *COLING 1998: The 17th International Conference on Computational Linguistics*, 519–525. ACL: Université de Montreal.

von Heusinger, Klaus. 2011. Specificity. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, 1025–1058. Berlin: Mouton De Gruyter.

von Heusinger, Klaus. 2019. Indefiniteness and specificity. In Jeanette Gundel & Barbara Abbott (eds.), *The Oxford handbook of reference*, 146–167. Oxford: Oxford University Press.

von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19(3). 245–274.

Himmelmann, Nikolaus. 2001. Articles. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, 831–841. Berlin: De Gruyter.

Kamp, Hans. 1981. A theory of truth and semantic interpretation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Matematisch Centrum.

Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and Discourse Representation Theory.* Dordrecht: Kluwer Academic Publishers.

Karttunen, Lauri. 1968. What do referential indices refer to? Paper prepared for the Linguistics Colloquium. University of California, Los Angeles.

Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1). 339–359.

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

Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.

Lyon, John. 2015. Okanagan determiner phrases and domain restriction. *International Journal of American Linguistics* 81(2). 187–219.

Maclaran, Rose. 1982. *The semantics and pragmatics of the English demonstratives.* Ithaca: Cornell University. (Doctoral dissertation).

McCawley, James D. 1979. Presupposition and discourse structure. In Choon-Kyu Oh & David A. Dinneen (eds.), *Syntax and semantics*, vol. 11: Presupposition, 371–388. New York: Acedmic Press.

Neale, Stephen. 1990. *Descriptions.* Cambridge, MA: MIT Press.

Peng, Danling. 2004. Cognitive processing of the Chinese language and its neural mechanism. *Contemporary Linguistics* 2004-04.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland & Alice G. B. ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, MA: MIT Press.

Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493.

Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. In Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages* (Studies in Diversity Linguistics 25), 1–37. Berlin: Language Science Press.

Simpson, Andrew, Hooi Ling Soh & Hiroki Nomoto. 2011. Bare classifiers and definiteness: A cross-linguistic investigation. *Studies in Language* 35(1). 168–193.

Strawson, Peter F. 1950. On referring. *Mind* 59. 320–344.

Tawa, Wako. 1993. Interpretation of definiteness: With special reference to Japanese. *Word* 44(3). 379–396.

Chapter 2

(In)definiteness and Vietnamese classifiers

Walter Bisang

Johannes Gutenberg University, Mainz & Zhejiang University

Kim Ngoc Quang

Johannes Gutenberg University, Mainz & University of Social Sciences and Humanities, Ho Chi Minh City

Vietnamese numeral classifiers (CL) in the bare classifier construction [CL+N] can be interpreted as definite and as indefinite. Based on a corpus of written and oral texts with a broad range of different contexts for the potential use of classifiers, this paper aims at a better understanding of the factors and linguistic contexts which determine the use of the classifier in [CL+N] and its specific functions. The following results will be presented: (a) Even though classifiers tend to be interpreted as definite, they are also used indefinitely, irrespective of word order (subject/preverbal or object/postverbal). (b) There is a strong tendency to use the [CL+N] construction with definite animate nouns in the subject position, while bare nouns [N] preferably occur with indefinite inanimate nouns in the object position. (c) The vast majority of nouns occurring with a classifier are sortal nouns with the features [−unique, −relational]. (d) Discourse and information structure are the most prominent factors which determine the grammar of Vietnamese classifiers. The influence of discourse is reflected in the pragmatic definiteness expressed by the classifier. Moreover, information structure enhances the use of a classifier in contexts of contrastive topic, contrastive focus and focus particles. Finally, thetic statements and some special constructions (existential clauses, verbs and situations of appearance) provide the environment for the indefinite interpretation of classifiers.

Walter Bisang & Kim Ngoc Quang

1 Introduction

Numeral classifiers are an areal characteristic of East and mainland Southeast Asian languages in the context of counting. This fact is well known and has been frequently discussed in the literature since the 1970s (Greenberg 1972). What is less well known and has been discussed only in more recent times is the use of the same classifiers in the contexts of definiteness and indefiniteness, when they occur in the [CL+N] construction (bare classifier construction; cf. Bisang 1999; Cheng & Sybesma 1999; Simpson 2005; Wu & Bodomo 2009; Li & Bisang 2012; Jiang 2015; Simpson 2017; Bisang & Wu 2017). In Vietnamese, classifiers in [CL+N] are clearly associated with reference. What is controversial in the litera-ture is the question of whether they are used only in the context of definiteness or in contexts of definiteness and indefiniteness. Tran (2011) claims that classi-fiers only have a definite interpretation, while Nguyen (2004) argues for both interpretations (see also Trinh 2011). A look at an example from Nguyen (2004) in (1) shows that both interpretations are possible. In this respect, it differs sig-nificantly from many Sinitic languages with [CL+N] constructions. While the definiteness/indefiniteness interpretation of classifiers depends on the preverbal or postverbal position of the [CL+N] construction in most of these languages,[1] Vietnamese classifiers can have both interpretations in both positions. In (1a), *con bò* [CL cow] is in the subject position and is open to both interpretations ('the cow'/'a cow'). Similarly, *cuốn sách* [CL book] in the object position of (1b) can be definite as well as indefinite ('the book'/'a book'):

(1) Nguyen (2004):
    a. Con bò ăn lúa   kìa!
       CL   cow eat paddy SFP
       'Look! A/the cow is eating your paddy!'
    b. Mang cuốn sách ra   đây!
       bring CL   book out here
       'Get a/the book.'

As can be seen from the following example, nouns without a classifier (bare nouns) can also be interpreted in both ways in both positions. In Nguyen's (2004) analysis, the only difference between the bare noun construction and the

---

[1] In Wang's (2015) survey of Sinitic classifiers as markers of reference, the definiteness/indefi-niteness distinction is independent of word order relative to the verb in only 10 out of his 120 sample languages (cf. Type I classifiers in his terminology).

[CL+N] construction is that the former can be interpreted as singular or plural, while the latter can only have a singular reading:

(2) Nguyen (2004):

    a. Bò   ăn lúa   kìa!
       cow eat paddy SFP
       'Look! A/the cow(s) is/are eating your paddy!'

    b. Mang sách ra   đây!
       bring book out here
       'Get a/the book(s), will you?'

Even though these examples show that classifier use is not obligatory and that classifiers can be interpreted as definite as well as indefinite, neither the conditions under which classifiers have these functions nor their specific referential meaning are well understood. Analyzing Vietnamese classifiers in the [CL+N] construction as variables whose interpretation depends on semantic, syntactic and discourse-pragmatic contexts, it is the aim of this paper to define the contexts which determine their use in terms of obligatoriness and their interpretation as definite and indefinite. Since the use of the classifier in the [CL+N] construction and its interpretation in terms of (in)definiteness in Vietnamese strongly depends on discourse and information structure, as in many other East and mainland Southeast Asian languages, looking at individual examples in isolation is not sufficient for modeling the function and the use of the [CL+N] construction. What is needed are texts, both written and oral. For that reason, we decided to set up our own corpus of Vietnamese, which is based on written and oral reports, by native speakers of Vietnamese, on the content of two silent movies (for details, cf. section 2).

The analysis of the data from our Vietnamese corpus confirms the general observation that classifiers in [CL+N] can be interpreted as definite as well as indefinite, irrespective of word order. It also shows that the interpretation of numeral classifiers in terms of definiteness and indefiniteness in [CL+N] depends on semantic and syntactic (preverbal/postverbal or subject/object) factors, as well as on discourse and information structure. In addition to that, it turns out that the definite function is much more frequent than the indefinite function. Instances of indefinite [CL+N] constructions mainly occur in special contexts and constructions, such as thetic statements, existential clauses, and constructions with verbs which introduce previously unidentified referents into discourse (i. e., verbs of appearance). Given the relative rareness and the functional specifics of indefinite

classifiers, it may not come as a surprise that the indefinite interpretations of classifiers remained unnoticed in a number of studies of Vietnamese classifiers.

To find out more about the function of classifiers and the factors that determine their use, the following criteria will be studied in more detail:

- Definiteness and indefiniteness of the nominal expression
- The semantic feature of [±animate] of the noun
- The semantic features of [±unique] and [±relational] in terms of Löbner's (1985; 2011) four basic types of nouns
- The syntactic criterion of word order (position of the noun in the subject/ preverbal or object/postverbal position)
- The role of discourse and the relevance of identifiability
- Information structure and the use of a classifier (contrastive topics, certain types of focus) as well as its function (i. e., theticity and indefiniteness)
- The combination with specific verbs (i. e., existential verbs and verbs of appearance).

The structure of the paper is as follows: after the discussion of methodological issues in section 2, section 3 will describe classifiers in their definite functions and the criteria that determine their use. Section 4 will do the same with classifiers in their indefinite function. The conclusion in section 5 will briefly summarize our findings and situate them with regard to other languages with numeral classifiers that are used in contexts of definiteness as well as indefiniteness.

2 Methodology

Our analysis of the function and the use of classifiers in the [CL+N] construction is based on a Vietnamese corpus of 30 written texts and 30 oral texts produced by native speakers of Vietnamese who were asked to report on the content of two films which were previously presented to them on the screen of a personal computer. One of the films was used to create a written corpus, the other an oral corpus. The total number of informants involved was 46 (25 female and 21 male informants). Fourteen informants (five female and nine male) from among these 46 informants participated in both experiments and thus produced a written and an oral text.[2] In total, there were 15 male and 15 female informants, as well as 15

---

[2]Since it was more difficult to find male informants, we had to ask more males to take part in both experiments. Six of the remaining 12 male informants only produced a written text, while the other six only were involved in the oral experiment.

graduate and 15 undergraduate informants for each experiment. The reason for this arrangement was to check for potential effects originating from differences in gender or modality (written vs. oral). Since we did not find any significant differences, we will not address this issue in the present paper.

The experiments were carried out by Kim Ngoc Quang in Ho Chi Minh city (Southern Vietnam) with the support of assistants who played the role of addressees (readers/listeners). This arrangement was necessary to avoid speaker assumptions about information shared with the addressee. Thus, the informants reported their stories in a situation in which it was clear that the addressee did not know the story.

For the purpose of our study, we needed two films with multiple protagonists, frequently changing scenes with different perspectives and a large number of animate and inanimate objects involved in a variety of actions expressed by transitive and intransitive verbs. The first film, with the title 'Cook, Papa, Cook', is a silent movie of nine minutes and 38 seconds in length.[3] This very lively film, which was used to create the written corpus, has three protagonists: a husband, a wife and their son. The story is characterized by intense quarrels between the husband and his wife. Because of this, the wife decides that she is no longer prepared to make breakfast for her husband. His attempts to make it himself are met by a number of obstacles and end up turning the kitchen into a total mess. When he finally manages to make his own kind of breakfast, his wife refuses to eat it.

The second film, which was used to set up the oral corpus, is from the 'Pear Stories' (Chafe 1980).[4] It is five minutes and 54 seconds long. It has two protagonists: a farmer and a young boy, who steals the farmer's pears from some baskets, while the farmer is up a tree picking the rest of the pears. When cycling away from the farmer, he inadvertently rides over a stone because he is distracted by a girl cycling in the opposite direction. As a consequence, the pears roll out of the basket and scatter all over the road. Three other boys arrive and help the boy to pick up the pears. As a reward for their help, the boy offers them each a pear. Later on, the three boys walk past the farmer while eating their pears. The film ends with the farmer trying to understand what has happened.

The length of the 30 written texts varies between 491 and 1,944 words. The written corpus as a whole consists of 31,663 words. The total length of the oral corpus is 17,777 words, after transcription. The length of the 30 oral texts varies between 321 and 1,061 words.

In this paper, the two corpora are employed as sources of examples of a

---

[3]The film can be seen on YouTube at `https://www.youtube.com/watch?v=OITJxh51z3Q`.

[4]The film can be seen on YouTube at `https://www.youtube.com/watch?v=bRNSTxTpG7U`.

broad range of different classifier functions and different conditions responsible for their occurrence. Moreover, the data from these corpora are used for some generalizations about frequency, as far as that is possible on the basis of calculating simple percentages.

3 Classifiers and definiteness

This section examines the correlation between classifiers in the [CL+N] construction[5] with a definite interpretation from various perspectives. Subsection 3.1 discusses the semantic feature of animacy and its interaction with definiteness. An examination of the semantic features of uniqueness and relationality in subsection 3.2 shows that the vast majority of nouns occurring with a classifier are sortal nouns, defined by their features of [−unique]/[−relational]. The interaction of word order (preverbal/subject and postverbal/object) with animacy and definiteness is explored in subsection 3.3. Finally, the roles of discourse (identifiability) and information structure (contrastive topics, focus particles and contrastive focus) are discussed in subsection 3.4.

3.1 Animacy and definiteness

Animacy plays an important role in grammar. This can be clearly seen from the animacy hierarchy as introduced by Silverstein (1976) and Dixon (1979), which is involved in such divergent domains of grammar as alignment, differential object marking, direct/inverse marking and number marking on nouns (to name just a few). An examination of this hierarchy in its full form, as it is presented in Croft (2003: 130), shows that it is not only concerned with animacy but also with person and referentiality.

(3) Animacy hierarchy (Croft 2003: 130):
    first/second person pronouns > third person pronoun > proper names > human common noun > non-human animate common noun > inanimate common noun.

The role of animacy in a strict sense is limited to the animacy scale, which goes from human to animate to inanimate. Animacy generally contributes to prominence (for a good survey, cf. Bornkessel-Schlesewsky & Schlesewsky 2009). Another important scale that contributes to prominence is the definiteness scale

---

[5]Notice that we do not discuss instances of [NUM CL N] with numerals > 1 because we do not have enough data in our corpus.

that runs from personal pronoun to proper name, to definite NP, to indefinite specific NP, to non-specific NP (cf. Aissen 2003, on the relevance of these two scales for differential object marking). As will be shown in this subsection, based on the Vietnamese data from our experiments, both scales have their impact in the use of classifiers inasmuch as there is a strong tendency for classifiers to be used with definite animate nouns.

As for animacy, Table 1 below shows a clear correlation between the feature of [±animate] and classifier use. Out of 1,698 instances with animate nouns, 1,571 instances[6] (92.5%) take a classifier, while only 127 instances[7] (7.5%) occur without a classifier. In contrast, only 742 instances[8] of [−animate] nouns (27.6%) occur with a classifier, while 1,948 instances[9] (72.4%) are bare nouns.[10]

Table 1: Token frequency of classifier use with [±animate] nouns in written texts and oral texts (in our Vietnamese corpus)

Nouns in narratives	[+animate]	[−animate]
[CL+N]	1,571 instances (92.5%)	742 instances (27.6%)
[N]	127 instances (7.5%)	1,948 instances (72.4%)

Our Vietnamese data also show that classifiers can be interpreted as definite as well as indefinite but that there is a strong tendency towards definite interpretation in our written and in our oral corpus. This can be seen from Table 2, in which 1,444 instances of [CL+N] in the written corpus are definite (92.0%; 1,154 + 290), while only 125 instances are indefinite (8.0%; 22 + 103). Similarly, the oral corpus shows 680 instances of classifiers in their definite function (91.4%; 395 + 285), which contrast with only 64 classifiers with an indefinite reading (8.6%; 0 + 64). The same table additionally shows that definiteness clusters with

---

[6]1,571 is the result of all [+animate] nouns with a classifier in the written corpus (978 + 19 + 176 + 3) plus all [+animate] nouns with a classifier in the oral corpus (262 + 0 + 133 + 0) in Table 6.

[7]127 is the result of all [+animate] nouns with no classifier in the written corpus (8 + 1 + 114 + 0) plus all [+animate] nouns with no classifier in the oral corpus (0 + 1 + 3 + 0) in Table 6.

[8]742 is the result of all [−animate] nouns with a classifier in the written corpus (34 + 9 + 256 + 94) plus all [−animate] nouns with a classifier in the oral corpus (55 + 2 + 230 + 62) in Table 6.

[9]1,948 is the result of all [−animate] nouns with no classifier in the written corpus (78 + 31 + 1,092 + 365) plus all [−animate] nouns with no classifier in the oral corpus (12 + 0 + 324 + 46) in Table 6.

[10]The frequencies of classifier use in the tables in this paper are for those occurrences in [CL+N] constructions; hence sequences such as *hai cuốn sách* [two CL book] 'two books' would not be counted in these tables.

animacy. In the written corpus, 1,154 animate definite nouns with a classifier (90.4%) contrast with only 122 animate definite nouns with no classifier (9.6%). In the case of oral texts, animate definite nouns reach an even higher percentage: 100% of these nouns take a classifier. As for inanimate definite nouns, only 19.9% of the written corpus (290 out of a total of 1,460) and 45.9% of the oral corpus (285 out of 621) take a classifier.

Table 2: Token frequency of [±animate] nouns and their interpretation as definite and indefinite in written texts and oral texts (in our Vietnamese corpus)

	Written texts				Oral texts			
	[+animate]		[–animate]		[+animate]		[–animate]	
	[CL+N]	[N]	[CL+N]	[N]	[CL+N]	[N]	[CL+N]	[N]
Definite	1,154	122	290	1,170	395	3	285	336
Indefinite	22	1	103	396	0	1	64	46
Total	1,176	123	393	1,566	395	4	349	382

The following two examples illustrate the use of animate nouns with a classifier. In (4), the classifier occurs with one of the human protagonists of the story, who is clearly identifiable and definite at the point at which he is mentioned in that example. In example (5), the classifier is interpreted as indefinite. The animate noun *dê* 'goat' is introduced into the story.[11] As will be seen later in subsection 4.2, the co-occurrence with the copula verb *là* 'to be' is one of the typical contexts in which [CL+N] is interpreted as indefinite (cf. example 31):

(4) [+animate, +CL, +DEF] (Oral text 19, sentence 12)

Cậu bé thấy thế tặng mỗi người một trái lê …
CL boy see that present each person one CL pear
'The boy saw that he gave each of them one pear.'

(5) [+animate, +CL, –DEF] (Oral text 16, sentence 7)

Có một người dẫn con, con đó chắc là **con dê**, đi ngang qua.
have one person lead CL CL DEM maybe CL goat go pass over
'There was a man who led a, a, it may be a goat, passing by.'

---

[11]Notice, however, that in the continuation of this text, the goat is further specified as a *dê núi* [goat mountain] 'wild goat' and does not take a classifier. With this type of compound, classifiers are often omitted.

The following example shows how inanimate nouns tend to be realized as bare nouns, even if they are definite. The referents expressed by *thang* 'ladder' and *cây* 'tree' have already been mentioned but do not have classifier marking:[12]

(6) [–animate, –CL, +DEF] (Oral text 27, sentence 3)

Sau đó, ông ấy lại leo lên **thang** và leo lên
after that 3.SG again climb PREP ladder CONJ climb PREP

cây hái tiếp.
tree pluck continue

'After that, he [the farmer] climbed up the ladder and climbed onto the tree again to continue picking [pears].'

The comparatively less frequent combination of inanimate nouns with classifiers is illustrated by the following two examples:

(7) [–animate, +CL, +DEF] (Written text 2, sentence 8)

Cậu ta đã đặt **cái xô** ngay giữa bố và mẹ.
3.SG PERF place CL bucket right between father and mother

'He put the bucket right between his father and mother.'

(8) [–animate, +CL, –DEF] (Written text 24, sentence 14)

Lúc này, người đàn ông thức dậy, lấy **cái bình** rót nước vào ly,
time DEM CL man wake up take CL bottle pour water PREP glass

'At this time, the man woke up, he took a bottle and poured water into a glass,'

In (7), the inanimate noun *xô* 'bucket' was previously introduced into the scene by one of the protagonists (the boy). Given that the bucket is activated in the hearer's mind, the classifier marks definiteness in this example. In (8), the noun *bình* 'bottle' refers to a newly introduced concept. Thus, the classifier *cái* marks indefiniteness in this context.

The relationship between animacy/definiteness and word order (the position of the [CL+N] construction relative to the preverbal and postverbal positions) will be discussed in subsection 3.3.

---

[12] One of our reviewers asks if *thang* 'ladder' and *cây* 'tree' may be analyzed as instances of incorporation into the verb plus preposition. Given that both referents represented by these nouns can be clearly identified from their previous mention as individuated countable concepts in the text, such an analysis does not seem to be very likely.

Walter Bisang & Kim Ngoc Quang

3.2 The semantic features of uniqueness and relationality

The distinction between [±relational][13] nouns and [±unique][14] nouns as discussed by Löbner (1985; 2011) is of crucial importance for describing the use of classifiers in Vietnamese. The combination of these features with their two values yields the following four basic types of nouns, which correspond to four types of concepts or four logical types: sortal nouns ([−relational]/[−unique]; <e,t>), individual nouns ([−relational]/[+unique]; <e>), relational nouns ([+relational]/[−unique]; <e,<e,t>>) and functional nouns ([+relational]/[+unique]; <e,e>).

Table 3 presents our data on the presence or absence of classifiers in the context of Löbner's (1985; 2011) basic types of nouns. As can be seen, the vast majority of nouns occurring with a classifier are sortal nouns ([−unique]/[−relational]): out of a total of 2,313 nouns with a classifier, 2,309 (99.8%) belong to this type. Moreover, only three [+unique] nouns (marked in bold) out of 108 (2+83+1+22) take a classifier (2.8%), while 105 of them are realized as bare nouns (97.2%). In a similar way, relational nouns ([−unique]/[+relational]) have a strong tendency to occur without a classifier. Only one out of a total of 57 instances of this type (1.8%) takes a classifier.

Table 3: Token frequency of classifier with [±relational], [±unique] nouns in written texts and oral texts (in our Vietnamese corpus)

[±relational], [±unique] nouns	[+relational]				[−relational]			
	[+unique] functional		[−unique] relational		[+unique] individual		[−unique] sortal	
	[CL+N]	[N]	[CL+N]	[N]	[CL+N]	[N]	[CL+N]	[N]
Written texts	2	76	0	48	0	2	1,567	1,563
Oral texts	0	7	1	8	1	20	742	351
Total	2	83	**1**	56	**1**	22	**2,309**	1,914

From the four non-sortal nouns with a classifier, two are used in anaphoric situations. In example (9), the [+unique/+relational] noun *mông* 'buttocks' is first introduced into the story by a bare noun. The second time it is mentioned, the same noun occurs with the general classifier *cái*, its interpretation being definite

---

[13]Relational nouns have not only a referential argument, but also an additional relational argument (cf. the relational noun *daughter [of someone]* in contrast to the absolute noun *girl*).

[14]Unique nouns denote concepts which are uniquely determined in a given situation (e. g., *the sun, the pope*). Notice that the default use of uniqueness is singular definite. Plural, indefinite and quantificational uses require special marking.

because the object it denotes is now activated in the hearer's mind:

(9) (Written text 1, sentence 45)

Bị   nóng **mông**,   anh ta mở   vòi-nước xịt   mát cho **cái mông**,
PASS hot   buttock 3.SG   open water-tap spray cool for CL buttock

thì   lúc đó, bạn   anh ta chồm từ   ngoài cửa sổ vào hối
CONJ time DEM friend 3.SG   prance from outside window in   urge

anh ta nhanh-lên kẻo   trễ giờ.
3.SG   hurry-up   otherwise late

'[His] buttocks were burnt, he turned on the tap and sprayed cool water
onto the buttocks, at that time, his friend gesticulated from outside of the
window to urge him to hurry up as otherwise he would be late.'

A similar pattern is found in example (10) with the [−unique/+relational]
noun *chân* 'leg', which is expressed by a bare noun when it is first mentioned.
Later on, it is taken up together with the general classifier *cái* expressing defi-
niteness in this context:

(10) (Oral text 4, sentence 21)

Lê đổ   ra tung toé,   hình như nó   bị   đau **chân** nữa, thấy
pear pour out everywhere seems   3.SG PASS hurt leg more, see   3.SG

nó   sờ   sờ **cái chân**.
touch touch CL leg

'The pears rolled out everywhere, it seemed that his leg was hurt,
(because I saw) he touched [his] leg.'

In the other two instances of the [CL+N] construction with a non-sortal
noun, the use of the classifier is due to information structure (focus). For that rea-
son, the relevant examples will be discussed in subsection 3.4.3 (cf. (23) and (25)).

3.3 Word order, definiteness and animacy

In many Sinitic numeral classifier systems, the referential status associated with
the classifier in [CL+N] constructions depends on word order relative to the verb
(see Wang 2015 for a survey). The following examples in (11) and (12) from Li &
Bisang (2012) show how the preverbal subject position and the postverbal object
position are associated with definiteness and indefiniteness in Mandarin, in the
Wu dialect of Fuyang and in Cantonese.

Walter Bisang & Kim Ngoc Quang

While the [CL+N] construction in the subject position is ungrammatical in Mandarin Chinese (11a), it is interpreted in terms of definiteness in the Wu dialect of Fuyang (11b) and in Cantonese (11c).

(11) [CL+N] in the subject position (Li & Bisang 2012: 338)
Context: Where is the book?

   a. Mandarin:

     nà   běn shū, **(*ge) xuéshēng** mǎi-zǒu   le.
     that CL book CL   student   buy-away PF
     'The book, the student(s) has/have bought it.'

   b. Wu Chinese:

     pen cy   ke iaʔsn ma le tɕhi die.
     CL   book, CL student buy PFV go SFP
     'The book, the student bought (it).'

   c. Cantonese:

     bun syu, go hoksaang maai-jo la.
     CL   book CL student   buy-PFV SFP
     'The book, the student bought (it).'

In the object position, the classifier in [CL+N] is associated with indefiniteness in Mandarin (12a) and the Wu dialect of Fuyang (12b). In Cantonese, it goes with definiteness and indefiniteness (12c):

(12) [CL+N] in the object position (Li & Bisang 2012: 338-339)

   a. Mandarin:

     wǒ mǎi-le   **liàng chē**.
     I   buy-PFV CL   car
     'I bought a car.'

   b. Wu Chinese:

     Nge ma le   **bu tsʰotsʰi**.
     I   buy PFV CL car
     'I bought a car.'

   c. Cantonese:

     Keuih maai-zo **gaa ce**.
     he   sell-PFV CL car
     'I sold a car/the car.'

As can be seen from Table 4, the situation is different in Vietnamese. The [CL+N] construction occurs preverbally and postverbally and the classifier can be associated with definiteness as well as indefiniteness in both positions. A closer look reveals that the definite interpretation of the classifier is generally preferred. The overall percentage of definite [CL+N] constructions is 91.8% in contrast to only 8.2% of classifiers with an indefinite function.[15] The dominance of the definite interpretation is even stronger in the subject position (cf. the figures printed in bold). If the written and oral texts are combined, 1,329 out of 1,359 [CL+N] constructions, or 97.8%, are definite.[16] In the object position, the asymmetry between the definite and the indefinite interpretation is not as strong as in the subject position. In spite of this, the definite interpretation still clearly dominates, with 795 (432 + 363) instances (83.3%), compared with only 159 (97 + 62) instances (16.7%) with an indefinite interpretation.[17]

Table 4: Token frequency of the presence/absence of a classifier in subject and object positions in relation to definite vs. indefinite function (in our Vietnamese corpus)

Nouns in narrative	Written texts				Oral texts			
	Subject		Object		Subject		Object	
	[CL+N]	[N]	[CL+N]	[N]	[CL+N]	[N]	[CL+N]	[N]
Definite	**1012**	86	432	1206	**317**	12	363	327
Indefinite	**28**	32	97	365	**2**	1	62	46
Total	1040	118	529	1571	319	13	425	373

The two examples in (13) and (14) illustrate the definite function of the classifier in [CL+N]. In (13), *con lừa* [CL donkey] 'the donkey' is in the subject position. Because it is mentioned in the previous context, the classifier *con* has a definite reading. In (14), *cô vợ* [CL wife] 'the wife' is in the object position. Since it is mentioned in the preceding text, it is also interpreted as definite:

---

[15] The total number of definite [CL+N] constructions is 2,124 (1,012 + 432 + 317 + 363); the total number of indefinite [CL+N] constructions is 189 (28 + 97 + 2 + 62).

[16] The total number of definite [CL+N] constructions in subject position is 1,329 (1,012 + 317); the total number of indefinite [CL+N] constructions is 30 (28 + 2).

[17] Recall that bare nouns in Vietnamese can also occur in both subject positions and object positions and be interpreted as either definite or indefinite.

Walter Bisang & Kim Ngoc Quang

(13) Definite subject (Oral text 26, sentence 9)

Con lừa   cứ   nhìn vào   các
CL   donkey always look inside PL

cần xé lê như muốn đứng lại và   ăn lê.
CL pear like want stop   CONJ eat pear

'The donkey kept on looking into the baskets as if it wanted to stand by and eat them.'

(14) Definite object (Written text 9, sentence 14)

Bực mình, anh chồng   đóng sầm cửa khiến **cô vợ**   giật mình,
angry   CL husband slam   door cause CL wife startled

rồi   bỏ   vào   nhà tắm.
CONJ leave enter bathroom

'Annoyed, the husband slammed the door. This upset [his] wife, then he went to the bathroom.'

The following two examples focus on the object position and indefiniteness (for indefinite [CL+N] constructions in the subject position, cf. subsection 4.1). At the same time, they also illustrate how classifiers in the same syntactic position can be interpreted as indefinite or definite, depending on context. In example (15) from our data on written texts, we find the same expression (*chiếc xe* [CL car] 'a/the car') in both functions.

(15) [Indefinite object, ±DEF] (Written text 1, sentence 26)

Anh ta bước vào nhà   thì   lại   bị   đứa con chơi **chiếc xe**$_1$
3.SG   step PREP house CONJ EMPH PASS CL   son play CL   car

đẩy   trúng vào   chân khiến anh ta ngã ngửa vào   **chiếc xe**$_2$.
push RES   PREP leg   cause 3.SG fall.back PREP CL   car

'When he entered the house, he ran into his son who was playing and he got hit by a car [a toy car] into [one of his legs]. [This] made him fall down onto the car.'

In the first line, the noun *xe* 'car' in *chiếc xe* is not activated by previous context. Thus, the classifier must be interpreted as indefinite. In the second line, the same car is taken up again with the same classifier (*chiếc*), which now has a definite interpretation. The next example is from our oral corpus:

(16) [Indefinite object, ±DEF] (Oral text 26, sentence 1)

Có   một người đàn ông đang ở   trên **cái thang** bắc   lên **cây**
exist one CL   man   PROG PREP top CL ladder connect PREP CL

lê và   đang hái   **trái lê**.
pear CONJ PROG pluck CL   pear

'There was a man on [a] ladder which was propped up against [a] pear
tree. He was picking [its] pears.'

In this example, we find three [CL+N] constructions, i. e., *cái thang* [CL$_{general}$ ladder], *cây lê* [CL$_{tree}$ pear] and *trái lê* [CL$_{fruit}$ pear]. Since the first two nominal concepts are newly introduced, the corresponding [CL+N] constructions are interpreted as indefinite ('a ladder' and 'a pear tree'). The third [CL+N] construction is associated with the previously mentioned pear tree. For that reason, the classifier *trái* for fruits can be interpreted as definite through bridging ('its pears [i. e. the pears of the previously mentioned tree]').

If the data on classifier use in the subject and in the object position is combined with the semantic feature of animacy as in Table 5, it can be seen that there is a clear preference for animate nouns in the subject position. There are 1,269 instances (85.2%) of [+animate] nouns in the subject position, which contrast with only 221 instances (14.8%) of [−animate] nouns. Similarly, the object position is characterized by its clear preference for [−animate] nouns. There are 2,469 [−animate] object nouns (85.2%) and only 429 [+animate] object nouns (14.8%). Thus, the data in Table 5 reflect the well-known preference of animate subjects and inanimate objects (cf. Givón 1979, Du Bois 1987 and many later publications).

Table 5: Distribution of instances of [±animate] nouns in the positions of subject and object (in our Vietnamese corpus)

	Subject		Object	
	[+animate]	**[−animate]**	**[+animate]**	**[−animate]**
Written texts	1,006	152	293	1,807
Oral texts	263	69	136	662
Total	1,269 (85.2%)	221 (14.8%)	429 (14.8%)	2,469 (85.2%)

Finally, the combination of the three parameters of word order (subject vs. object), reference (definite vs. indefinite) and animacy (animate vs. inanimate) yields the following results for the presence/absence of the classifier ([CL+N] vs. [N]):

Walter Bisang & Kim Ngoc Quang

Table 6: Presence/absence of classifiers depending on the features of [±animate], subject vs. object and definite vs. indefinite (in our Vietnamese corpus)

[+def] vs. [−def]	Written texts							
	Subject				Object			
	[CL+N] [±ani]		[N] [±ani]		[CL+N] [±ani]		[N] [±ani]	
	+	−	+	−	+	−	+	−
+def	978	34	8	78	176	256	114	1,092
−def	19	9	1	31	3	94	0	365
Total	997	43	9	109	179	350	114	1,457

[+def] vs. [−def]	Oral texts							
	Subject				Object			
	[CL+N] [±ani]		[N] [±ani]		[CL+N] [±ani]		[N] [±ani]	
	+	−	+	−	+	−	+	−
+def	262	55	0	12	133	230	3	324
−def	0	2	1	0	0	62	0	46
Total	262	57	1	12	133	292	3	370

Table 6 reveals that, of the 1,012 definite [CL+N] constructions in the subject position of the written text corpus, 978 (96.6%) are [+animate] nouns. Only 34 definite [CL+N] constructions in the subject position (3.4%) are [−animate]. Similarly in oral texts, 262 animate definite subject [CL+N] constructions (82.6%) contrast with only 55 inanimate definite subject [CL+N] constructions (17.4%). In the object position, the percentage of animate nouns with definite subject [CL+N] constructions is much lower: 40.7% (176 vs. 256) in the corpus of written texts and 36.6% (133 vs. 230) in the corpus of oral texts. The results from Table 6 combined with the results from Table 4 (general preference of definite classifier interpretation, particularly with [CL+N] constructions in the subject position) plus Table 5 (preference of animate subjects) show that the classifier prototypically occurs with definite animate nouns in the subject position.

These observations can be visualized more clearly by means of the bar chart in Figure 1. The blue columns represent definiteness, while the green ones stand for indefiniteness:
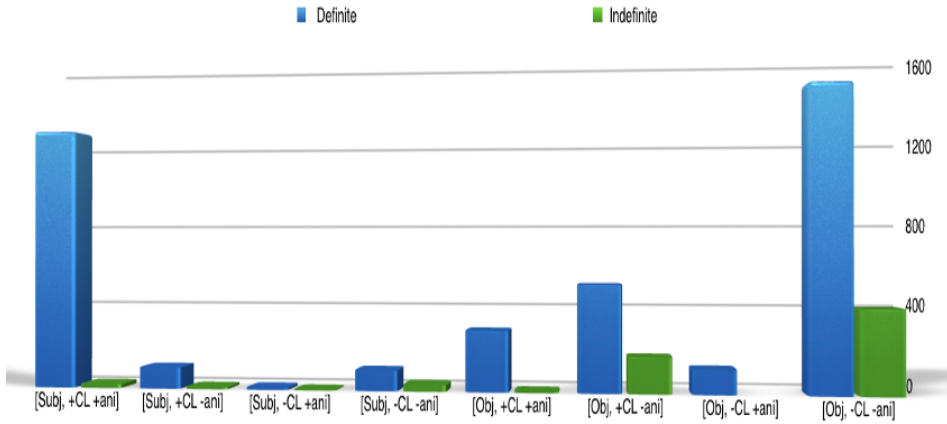
Figure 1: Token frequency of [±animate] nouns in subject and object function, marking definiteness or indefiniteness with or without a classifier (in our Vietnamese corpus)

In accordance with the data in Table 6, the blue columns representing definiteness are generally higher than the green columns, reflecting again the overall dominance of the definite function of Vietnamese classifiers. Moreover, the blue column in Figure 1 clearly dominates over the green column at the leftmost pole representing animate subjects with classifiers [Subj, +CL, +ani]. The preference for classifier use with animate subjects is further corroborated if the total number of tokens with the features [Subj, +CL, +ani] in the written and the oral corpus is compared with the total number of tokens with the features [Subj, −CL, +ani]. The figure for [Subj, +CL, +ani] is 1,259 (978 + 19 + 262 + 0), while the figure for [Subj, −CL, +ani] is just 10 (8 + 1 + 0 + 1). Thus, the use of the classifier with animate subjects overwhelmingly dominates over its absence with 99.2%. In addition to these results, the rightmost pole in Figure 1 with the features [Obj, −CL, −ani] demonstrates that inanimate object nouns tend to occur without a classifier. The overall number of tokens with the features [Obj, −CL, −ani] from the written and the oral texts is 1,827 (1,092 + 365 + 324 + 46), while the overall number of tokens with the features [Obj, +CL, −ani] is only 642 (256 + 94 + 230 + 62). Thus, the percentage of inanimate object nouns without a classifier is 74.0% against 26.0% with a classifier. Taken together, there is a clear preference for animate subjects to occur with a classifier and for inanimate objects to occur as bare nouns.

To conclude, the data presented in this subsection show that the (in)definiteness interpretation of the classifier is not rigidly determined by the position of the [CL+N] construction relative to the verb (subject vs. object position). In

fact, there is an overall preference for interpreting classifiers in [CL+N] as definite even though indefinite [CL+N] constructions are found in both positions. In spite of this, there are other factors which operate against this general tendency as well as against the use of classifiers in definite contexts. The semantic factors were presented above in subsection 3.2. Subsection 3.4 will discuss aspects of discourse and information structure.

3.4 Discourse and information structure

Discourse and information structure affect the meaning of Vietnamese classifiers as well as their presence or absence in a given context. As discussed in subsection 3.4.1 on meaning, the definiteness expressed by the classifier is discourse-based. The same subsection also shows how discourse enhances the use of classifiers with [+unique] nouns which otherwise show a strong preference for occurring as bare nouns in our data (cf. subsection 3.2). Subsections 3.4.2 and 3.4.3 illustrate how information structure determines the presence of a classifier. It will be shown that contrastive topics generally take a classifier (cf. subsection 3.4.2). Similarly, focus, as it manifests itself in contrastive focus and focus particles, can support the use of a classifier, even with non-sortal nouns (subsection 3.4.3).

3.4.1 Definiteness, identifiability and information structure

Classifiers in [CL+N] constructions very rarely occur with [+unique] nouns (cf. subsection 3.2 on the strong preference for sortal nouns ([−unique]/[−relational]). Moreover, the majority of definite classifiers are used in anaphoric contexts, in which a previously introduced concept is taken up with a classifier in order to highlight the speaker's assumption that it can be identified by the hearer (cf. examples (4), (7), (13), (14) and (15)). Even two of the four non-sortal nouns with a classifier acquire their classifier in an anaphoric context (cf. (9) and (10); for the other two, cf. subsection 3.4.3 on focus). Taken together, these facts are strong indicators that the definiteness expressed by the classifier marks pragmatic definiteness rather than semantic definiteness in terms of Löbner (1985). In Schwarz's (2009; 2013) framework, Vietnamese definite classifiers express anaphoric or "strong" definiteness rather than unique or "weak" definiteness.

      With these properties, the definiteness associated with the classifier corresponds to the findings of Li & Bisang (2012: 17) on identifiability. As they show in example (17) from the Wu dialect of Fuyang, uniqueness is not a necessary condition for the definite interpretation of the [CL+N] construction. Unique concepts can be expressed either by bare nouns or by the [CL+N] construction. A

[+unique] [−relational] noun like *thin* 'sky' in (17) occurs in its bare form if the sky is understood generically as the one and only one sky. Thus (17a) is a generic sentence expressing the fact that the sky is blue in general. In contrast, the classifier in *ban thin* [CL sky] (17b) indicates that the speaker means the sky as it is relevant for a given speech situation with its temporal or spatial index, and that s/he thinks that the hearer can identify it (Li & Bisang 2012: 17):

(17) Wu dialect of Fuyang (Li & Bisang 2012: 17):
    a. Generic use:
       Thin zi lan ko.
       sky be blue SFP
       'The sky is blue (in general).'
    b. Episodic use:
       Ban **thin** gintsɔ man lan.
       CL_piece sky today very blue
       'The sky is blue today.'

In Vietnamese, the situation seems to be similar. Since a much larger corpus than the two corpora used here would be needed to find examples like (17), we present another example from a Vietnamese dictionary in (18) (Nguyen et al. 2005: 116 and 1686). In (18a), we find *trời* 'sky' as a bare noun. In this form, the sky is understood generically as the endless outer space seen from the earth with its general property of being full of stars. In contrast, *bầu trời* [CL sky] 'the sky' in (18b) with a classifier denotes the inner space seen from the earth as it is currently relevant to the speech situation. The speaker employs the classifier to inform the hearer that s/he is referring to the sky as it currently matters and as it can be identified by the speaker and the hearer in a shared temporal or spatial environment.

(18) a. Trời đầy sao.
       sky full star
       'The sky is full of stars.'
    b. **Bầu** trời đêm nay đầy sao.
       CL sky tonight full star
       'THE sky tonight is full of stars.'

Further evidence for the discourse-dependency of classifier use with [+unique] nouns comes from the fact that the noun *trời* 'sky' can take several different classifiers, e. g., *bầu trời* [CL_round sky], *khung trời* [CL_frame sky] or *vùng trời* [CL_area sky],

etc. The selection of a specific classifier out of a set of possible classifiers depends on the particular property of the sky the speaker wants to highlight to facilitate its identifiability to the hearer. In such a situation, selecting a particular classifier is even compulsory:

(19) *(**Khung/bầu/vùng**) trời mơ ước của   hai chúng ta đây rồi!
       CL   sky dream POSS two 2.PL   here SFP
      'Our dream sky/world is here!'

In the above example, the speaker creates a specific notion of the sky as it is relevant for her/him and the hearer. This 'dream sky' is then anchored in space and time as relevant to the speech situation by a classifier.

In another of our four examples of non-sortal nouns with a classifier in (23), the [+unique, −relational] noun *đất* 'earth, ground' is marked by the classifier *mặt* 'face/surface' in a situation of contrastive focus. As in the case of the sky in (19), this noun is also compatible with other classifiers, among them *mảnh/miếng* 'piece' and *vùng* 'area'. The selection of a specific classifier depends again on the properties of the concept expressed by the noun as they are relevant to the speech situation.

3.4.2 Contrastive topics

There is an impressive body of literature on contrastive topics. For the purpose of this paper, Lambrecht's (1994: 183, 291, 195) discourse-based definition in terms of two activated topic referents which are contrasted will be sufficient. This type of topic is quite frequent in our Vietnamese corpus. A look at the statistics shows that classifier use is very strongly associated with contrastiveness. In fact, there is a classifier in each of the 84 instances of contrastive focus (66 in the written corpus and 18 in the oral corpus). Moreover, all nouns occurring in this function are [+animate].

In most examples, the action/state of one protagonist is contrasted with the action/state of another protagonist. As shown in (20), the actions of the son in the kitchen are contrasted with the actions of his mother in the bedroom (described as 'the wife' from the perspective of the husband). The son takes the classifier *đứa* for young boys, while the mother takes the classifier *bà* for women. The contrast between these two protagonists is supported by the adverbial subordinator *còn* 'while/whereas':

(20)   (Written text 26, sentence 23)

Đứa con trai thì đứng lên kệ-bếp   và   vẽ   bậy   lên tường,
CL   son   TOP stand up kitchen-bar CONJ draw disorderly on wall

còn   **bà vợ**   thì nằm ăn   đồ ăn nhanh với vẻ mặt   khoái chí.
while CL wife TOP lie   eat fast.food   PRE expression delightful

'[His] son stood on the kitchen base (cabinet) and scribbled [something]
onto the wall, while [his] wife was lying in bed, eating fast food with a
facial expression of delight.'

In (21), the husband is contrasted with his wife. The husband's anger and
his intention to make his wife eat some food is mirrored against his wife's reac-
tion of refusing to give in. Both nouns take a classifier. The husband occurs with
the classifier *ông* for men and the wife again takes the classifier *bà* for women.
The contrast is explicitly expressed by the disjunctive conjunction *nhưng* 'but':

(21)   (Written text 26, sentence 36)

Thấy thái độ của vợ-mình, **ông chồng**   điên-máu-lên và
see   attitude POSS wife-self CL husband get.crazy   and

bắt ép ăn, **nhưng bà vợ**   vẫn không ăn.
force eat CONJ CL wife still NEG   eat

'Seeing the behaviour of his wife, the husband went crazy and [tried to]
force her to eat, but [his] wife still did not eat.'

In the final example of this subsection, there is a contrast between a pro-
tagonist and a non-protagonist. The noun *bé* 'boy', as one of the two protagonists
in the Pear Story, is contrasted with the children (*trẻ* 'child'). What is contrasted
is the boy's action of leaving on a bike and the children's action of walking away.
Again, both nouns occur with a classifier (*thằng* for the boy and *bọn* for the chil-
dren) and there is a contrastive conjunction (*còn* 'while, whereas'):

(22)   (Oral text 6, sentence 31)

Thằng bé tập tễnh dắt xe   đi vài bước, còn **bọn trẻ** thì
CL   boy limping lead bike go few step CONJ CL kid TOP

đi theo   hướng   ngược lại.
go toward direction opposite

'The boy led the bike limpingly, while the children walked in the opposite
direction of the boy.'

Walter Bisang & Kim Ngoc Quang

3.4.3 Focus

Classifiers are also selected in various types of focus. This will be shown by the discussion of the two remaining non-sortal nouns with a classifier (cf. subsection 3.2) plus two additional examples. The first example is on the [+unique, −relational] noun *đất* 'earth/ground'. In (23), this noun is interpreted as definite by the classifier *mặt*[18] for flat surfaces because it has the function of contrastive focus. The author of this text starts her story from the perspective of the protagonist, a farmer, who is up 'on a tree' (*trên một cái cây*). Having described a series of the farmer's actions up there, her attention suddenly moves to the position of the baskets 'down on the ground' (*dưới mặt đất*), which is contrasted to the position up on the tree.[19]

(23) (Oral text 13, sentence 8)

Ổng leo   lên một cái thang để   ổng leo   lên một cái cây để
3.sg climb PREP one CL ladder so that 3.sg climb PREP one CL tree to

ổng hái. Ổng hái xong, thì ổng leo   xuống cái thang đó,
3.sg pluck 3.sg pluck RES CONJ 3.sg climb down CL ladder DEM

xuống đó. Rồi ổng, **dưới mặt đất**   sẽ có ba   cái giỏ   ...
down DEM CONJ 3.sg down CL   ground FUT have three CL basket
'He climbed a ladder to get on [a] tree to pick [the fruits]. Having picked [them], he went down [the] ladder. Then, he, down on the ground, there were three baskets...'

In contrast to (23), *đất* 'earth, ground' does not have a classifier in the non-contrastive situation of the following example:

---

[18] *Mặt* has the meaning of 'face'. In this context, it is a classifier for objects with a flat surface. As a full noun, it can be interpreted as a [+relational] noun as in *mặt bàn* [surface table] 'the surface of the table'.

[19] One of our reviewers suggests that *dưới mặt đất* 'down on the ground' is a frame-setter (e. g., Krifka 2008). This interpretation cannot be fully excluded. However, we would like to point out that the contrast between the position 'up in the tree' and the position 'down on the ground' is clearly given in the way the scenes are presented in the film.

(24)   (Oral text 28, sentence 20)

Thì   có   ba   sọt trái cây dưới **đất**,
CONJ have three CL fruit   under ground

không có   ai   trông nom hết.
NEG   have who take-care at.all

'There were three baskets of fruit on the ground, but nobody was taking care of them.'

Another context that induces classifier use is the context of focus particles, which typically marks the inclusion or exclusion of alternatives (König 1991). The other two examples to be discussed here both belong to this type of focus. The first example (25) is on the [+unique/+relational] noun *mặt* 'face', which occurs with the two focus particles *chỉ còn* 'only' and *mỗi* 'only'. The noun *mặt* 'face' takes the position between these two particles to emphasize the fact that the foam covers almost the whole of the husband's body, leaving only his face unaffected. Thus, the two particles exhaustively single out one part of the body, which is excluded from the disturbing presence of foam:

(25)   (Written text 29, sentence 31)

Lúc bấy giờ, người chồng   nghe thấy bèn   trồi lên khỏi mặt
time that   CL   husband hear RES   CONJ rise out out of surface

nước, toàn   thân ông là   bọt xà phòng chỉ còn thấy mỗi **khuôn mặt**.
water whole body 3.SG COP foam   only   see only CL   face

'At that time, the husband heard (the bell), then he moved out of the water. His whole body was full of soap foam, except **[the] face** [lit.: one can just only see **[his] face**].'

Our next two examples are not included in the statistics in Table 3 because they contain a possessive construction, and thus go beyond the distinction of bare noun vs. [CL+N]. In spite of this, they are relevant because classifiers very rarely occur with non-sortal head nouns of possessor constructions. In (26), the [+unique, +relational] possessee head noun *chồng* 'husband' in *chồng của mình* [husband CL self] 'husband of her' takes the classifier *ông*. Since non-sortal nouns of this type do not have a classifier in our data (e. g., *chồng (của) mình* [husband (possessive marker) self-reflexive pronoun] '[her] husband', *vợ (của) mình* [wife (possessive marker) self-reflexive pronoun] '[his] wife', *con trai họ* [son (possessive marker) selves-reflexive pronoun] '[their] son', etc.), it is reasonable to assume that the presence of the classifier is due to the focus particle *ngoài* 'except':

Walter Bisang & Kim Ngoc Quang

(26)   (Written text 30, sentence 8)

Khi   giật mình vì   bị   tạt   nước, bà vợ liền
when startle   because PASS throw water, CL wife immediately

thức giấc nhìn xung quanh xem ai   làm và   chả có   ai   **ngoài ông**
awake   look around   see who do   CONJ NEG have who except CL

chồng   của mình.
husband POSS self

'Being startled by the water, the wife awoke immediately, looked around
to see who did it. But there was nobody, except [her] husband.'

Finally, the classifier even occurs with non-sortal head nouns of posses-
sive constructions, if the relevant focus situation can only be derived from con-
text without the explicit presence of a focus marker. This is illustrated by (27),
in which we find the two non-sortal nouns *chân* 'foot' and *mông* 'buttocks', the
former without a classifier, the latter with a classifier. The interpretation of this
sentence crucially depends on the function of the adverbial subordinator *nên*
'therefore, to the extent that' which creates a context in which the situation be-
comes worse and worse until it culminates in a rather unexpected situation, in
which the husband even burns his buttocks. This situation can be compared to
the situation created by a focus particle like *even*:

(27)   (Written text 13, sentence 35)

Ông ấy bị   bỏng và   đau quá nên ôm **chân** lên và   không giữ
3.SG   PASS burn CONJ hurt very CONJ hold foot RES and NEG   keep

được thăng bằng nên té   vào chiếc chảo đang cháy đó, **cái mông**
RES balance   CONJ fall into CL   pan PROG burn DEM CL buttock

của ông ấy đã   bị   phỏng.
POSS 3.SG   PERF PASS burn.

'He got burnt and he got hurt, therefore, he lifted [his] leg up to hold it,
then he was no longer able to keep his balance to the extent that he fell
down on [the] burning pan and [as a consequence] even [his] buttocks
got burnt.'

4 Classifiers and indefiniteness

Classifiers with indefinite interpretation are limited to particular contexts: the
indefinite function of classifiers in the subject position of thetic statements is

presented in subsection 4.1. Subsection 4.2 discusses the [CL+N] construction in existential clauses, while subsection 4.3 describes [CL+N] constructions in combination with verbs of appearance.

4.1 Thetic statements

As can be seen from Table 4, indefinite subjects are rather rare: 97.8% of the preverbal [CL+N] constructions of the written and the spoken corpus together are definite (cf. subsection 3.3). The vast majority of the remaining 2.2% of indefinite preverbal [CL+N] constructions are subjects of thetic constructions (Kuroda 1972; Sasse 1987; 1995). Thetic utterances are seen in contrast to categorical utterances. Sasse (1995) defines both types as follows:

> Categorical utterances are said to be bipartite predications, involving a **predication base**, the entity about which the predication is made, and a **predicate**, which says something about the predication base. In other words, one of the arguments of the predicate is picked out as a "topic" in the literal sense, namely, an object about which something is asserted. Thetic utterances, on the other hand, are **monomial** predications (called "simple assertions" in Sasse 1987); no argument is picked out as a predication base; the entire situation, including all of its participants, is asserted as a unitary whole. (Sasse 1995: 4-5)

In utterances of this type, the entire clause is an 'all-new' utterance that is seen as inactivated information (often backgrounded) that is assumed by the speaker not to be present in the hearer's mind. Thus, nominal participants of thetic utterances are generally indefinite. The following two examples constitute the beginning of the story as told by two different informants. They provide a description of the initial scene as it was presented in the film. In the first sentence of both examples, the subject *đồng hồ báo thức* 'alarm clock' is marked by the classifier *chiếc*. Similarly, the subject *đàn ông* 'man' has the default classifier for humans *người* in the second sentence of both examples:

(28) Indefinite Subject (Written text 12, sentence 1, 2)

Chiếc đồng hồ báo thức reo lên lúc 8 giờ đúng.
CL clock alarm ring up at eight o'clock exactly

Người đàn ông đang ngủ thì bị nước văng tung tóe vào mặt.
CL man PROG sleep CONJ PASS water splatter PREP face

'The alarm clock rang at exactly eight o'clock. There was a man, who was sleeping and then [his] face was splattered with water.'

Walter Bisang & Kim Ngoc Quang

(29) Indefinite Subject (Written text 1, sentence 1, 2)

Chiếc đồng hồ báo thức reo lên báo hiệu đã tám giờ sáng.
CL clock alarm ring RES signaling PERF eight o'clock morning

Người đàn ông mở mắt liếc nhìn sang vợ-mình.
CL man open eye glance toward wife-self

'[The] alarm clock rang to signal that it was already 8 o'clock in the morning. [A] man opened his eye and glanced at [his] wife.'

4.2 Existential expressions

Existential sentences of the type 'there is an X' are typically used to introduce previously unidentified referents. Thus, [CL+N] constructions occurring in this type of construction are typically indefinite. Since they are positioned after the verb, they form a considerable part of the indefinite object classifiers in our data (but cf. inanimate nouns below). A good example is (30) from our oral corpus, in which the [CL+N] construction is preceded by the verb *có* 'have, there is':

(30) (Written text 20, sentence 28)

Lúc này, có **viên cảnh sát** vào hỏi xem tình hình vì hai
time DEM have CL policeman enter ask see situation because two

vợ chồng cãi nhau.
wife husband argue RECIP

'This time, [a] policeman entered and asked why this couple was arguing with each other.'

Another verb that implies indefiniteness is the copula verb *là* 'be', which is used in identificational contexts ('this is an X') as well as in locative contexts ('Y is [placed] in/at/on an X'). The following example starts out with a locative expression in the topic position (*bên cạnh đó* 'at the side of it, beside'). The three subsequent objects following the copula *là* are introduced as previously unmentioned elements into the scene by being situated within that locative topic:

(31) (Written text 14, sentence 2)

Bên cạnh đó là **cái kệ** **nhỏ**, **cái bình** và **ly nước** được đặt
beside   DEM COP CL shelf small CL bottle and CL water PASS place
lên   trên nó.
move.up top 3.SG

'Beside [him] was a small shelf with a bottle and a glass of water placed on it.'

Previous Context: The man who wore glasses awoke, opened his eyes for a moment, had a look around himself, ignored the alarm clock and went on sleeping.

In contrast to the thetic utterances of the preceding subsection, existential constructions can also be combined with constructions other than [CL+N]. For that reason, their impact on postverbal indefinite classifiers in our data is less strict than the impact of thetic utterances on indefinite classifiers in the subject position. As is shown by the following example, existential expressions can also occur with the [*một* 'one'+CL+N] construction:

(32) (Oral text 20, sentence 1, 2)

Câu chuyện được bắt đầu vào một buổi sang tại một cánh đồng,
CL story PASS begin in one morning at one CL field

có **một người nông dân** leo lên trên một cái thang, đang hái
have one CL farmer climb up on one CL ladder PROG pluck

một loại trái cây nào đó giống trái lê.
one kind fruit some certain like CL pear

'The story began in a morning in a field. There was a farmer, who was climbing up a ladder to pick a kind of fruit like a pear.'

Finally, there are also some instances of inanimate nouns which occur without a classifier in existential constructions. This is illustrated by the following example with the noun *xe cứu hoả* 'fire truck' in its bare form:

(33) (Written text 20, sentence 25)

Gần đó, có **xe cứu hỏa** và lập tức đến xịt nước vào
nearby have car save fire CONJ immediately arrive spray water into
chữa cháy nhưng làm cho mọi thứ hỏng hết.
extinguish fire CONJ cause everything ruin

'Nearby, there was [a] fire truck, it arrived immediately to extinguish the fire. However, it also ruined everything.'

Walter Bisang & Kim Ngoc Quang

The extent to which the use of the classifier ultimately depends on the animacy of the noun cannot be determined from our data because we do not have enough examples.[20]

4.3 Verbs and situations of appearance

Vietnamese has quite a few verbs with the meaning of 'appear, come up', 'turn out to be' or 'reveal', whose subsequent nouns introduce previously unidentified elements into the discourse. In such cases, the postverbal noun is indefinite. In the following example with the verb *lòi ra* 'come to light, appear', the noun *tẩu thuốc* 'smoking pipe' takes the general classifier *cái*. Since the pipe was hidden in the husband's pocket, it is unknown to the audience/reader of the text and is interpreted as indefinite.

(34) (Written text 15, sentence 61)

Nhưng sau đó, cái túi áo của ông chồng bị lủng, lòi ra **cái**
conj after that cl pocket poss cl husband pass burst, show out cl
tẩu thuốc, chứ không phải vật gì có thể gây nguy hiểm.
smoking-pipe conj_emph neg something can cause danger
'However, after that, his pocket burst and what came to light was a smoking pipe, definitely nothing that may cause any danger.'

Sometimes, the meanings of verbs implying the emergence of unidentifiable concepts are highly specific. This can be shown by the verb *vấp* 'trip, walk into, stumble over', which creates a situation in which the object is unpredictable and has the status of being unidentifiable as in the following example:

---

[20]In an alternative analysis, readers may be tempted to argue that the absence of the classifier is related to the complexity of the head noun (compounds vs. simple nouns) or to its status as a lexical item borrowed from Chinese. Since Emeneau (1951), it has often been claimed that nouns of this type take no classifiers. In spite of this, the noun *cảnh sát* 'policeman', which is borrowed from Chinese 警察 *jǐngchá* 'police(man)', does occur with the classifier *viên* in (30). Thus, we can at least exclude borrowing from Chinese as a strong factor for determining classifier use in existential constructions. In (31) it seems that animacy is more important. Ultimately, more data would be needed for to enable more precise conclusions to be reached.

(35) (Oral text 25, sentence 13)

Mải mê nhìn gái nên nó vấp phải
passionately look girl CONJ 3.SG trip PASS

cục đá và té xuống đường.
CL stone CONJ fall down road

'[He] looked at the girl passionately and thus stumbled over [a] stone and fell down on the road.'

Thus, the object *đá* 'stone' is marked by the classifier *cục* in (35). The boy, who is one of the two protagonists in the story, as well as the audience, cannot know what will happen when the boy is looking at the girl rather than at the road while riding his bike. The stone is clearly not activated and is interpreted as indefinite.

5 Conclusion

The aim of this study was to reach a better understanding of the referential functions of Vietnamese classifiers based on the systematic analysis of data from a corpus of written and oral texts which was designed to generate a broad variety of contexts which may trigger classifier use. The main results on the use and the functions of the Vietnamese classifier in [CL+N] can be summarized as follows:

(i) Classifiers can be interpreted as definite as well as indefinite but there is a clear preference for using the classifier in definite contexts (cf. subsection 3.1).

(ii) There is a clear clustering of animacy and definiteness: definite animate nouns occur much more frequently with a classifier than definite inanimate nouns (subsection 3.1).

(iii) There is a clear clustering of [CL+N] with [+definite, +animate, subject] and of bare nouns [N] with [−definite, −animate, object] (subsection 3.3).

(iv) The overwhelming majority of nouns occurring in the [CL+N] construction are sortal nouns [−unique, −relational] (subsection 3.2).

(v) Discourse and information structure play an important role in the function as well as in the presence/absence of a classifier:

  a. The definiteness with which classifiers are associated in [CL+N] is based on identifiability in discourse (subsection 3.4.1);

  b. Information structure is an important factor for determining the use of a classifier in [CL+N] (subsections 3.4.2 and 3.4.3) and its interpretation in terms of definiteness vs. indefiniteness (particularly cf. subsection 4.1 on indefiniteness and theticity).

(vi) There are certain semantic environments which support the indefinite interpretation of the classifier (existential clauses and verbs of appearance; subsections 4.2 and 4.3).

The results in (i) to (iii) on animacy, definiteness and subject/preverbal position tie in with general findings on prominence at the level of the morphosyntax-semantics interface as they manifest themselves in hierarchies like the animacy hierarchy (Silverstein 1976; Dixon 1979) or the accessibility hierarchy (Keenan & Comrie 1977) (for a survey, cf. Bornkessel-Schlesewsky & Schlesewsky 2009).[21]

---

[21]Based on the relevance of (in)definiteness and animacy, one may think of analyzing the use of the classifier in [CL+N] in the light of Differential Object Marking (DOM) as suggested by one of our reviewers. In our view, such an account would be problematic for at least the following reasons: (i) The use of the classifier in the [CL+N] construction is strongly associated with sortal nouns ([−relational]/[−unique]), while DOM marking is not limited to this type of noun. (ii) As is pointed out by Aissen (2003: 439), "it is those direct objects which most resemble typical subjects that get overtly case-marked". If one takes the use of the classifier as a DOM marker, one would expect the highest frequency of classifier use with [+definite] and [+animate] objects. This is clearly not borne out in the case of definiteness. As can be seen from Table 4, the ratio of definite subjects with CL is much higher than the ratio of definite objects with CL. There are 1,329 [= 1012 + 317] definite subjects with CL vs. 98 [86 + 12] definite subjects with no CL, i. e., 93.3% of the definite subjects in our two corpora have a classifier. In contrast, only 34.1% of the definite objects have a classifier (795 [= 432 + 363] definite objects with CL contrast with 1,533 [= 1206 + 327] without CL). In the case of animacy, the difference between the two ratios is smaller but it is still higher with animate subjects. As can be seen from Table 6, there are 1,259 [= 997 + 262] animate subjects with CL and only 10 [= 9 + 1] animate subjects with no CL, i. e., 99.2% of the animate subjects have a classifier. In the case of animate objects, the ratio is 72.7% (312 [= 179 + 133] animate objects with CL contrast with 117 [= 114 + 3] animate objects with no CL). (iii) The results discussed in (ii) are remarkable from the perspective of split vs. fluid DOM languages in terms of de De Hoop & Malchukov (2007). In split languages, DOM marking is obligatory for a particular feature, while it is optional in fluid systems. In most DOM languages, DOM is split for at least one category. As can be seen in (ii), this is not the case with the use of the classifier. Vietnamese classifiers are not obligatory with definite objects nor are they obligatory with animate objects.

The clustering observed in (ii) and (iii) additionally reflects a universal tendency to associate animate subjects in clause-initial positions of SVO languages with definiteness (Keenan & Comrie 1977; Givón 1979; Du Bois 1987; and many others). This tendency is also well known for word order in Sinitic languages (Li & Thompson 1976; Sun & Givón 1985; LaPolla 1995). Chen (2004: 1166) talks about definiteness-inclined preverbal positions and indefiniteness-inclined postverbal positions in Mandarin Chinese. As can be seen from (i), word order does not determine the (in)definiteness interpretation of the classifier in Vietnamese as rigidly as it does in Cantonese or in the Wu dialect of Fuyang (cf. the discussion of (11) and (12); for the discourse-based reasons for this, cf. below).[22]

The observation in (iv) that the vast majority of nouns occurring in the [CL+N] construction are sortal nouns in terms of Löbner (1985) confirms and further specifies the findings of Simpson (2017: 324) on the Wu variety of Jinyun, that nouns denoting "specifically unique individuals/elements" predominantly appear as bare nouns [N] (cf. the three instances of [+unique] nouns taking a classifier in Table 3). These results show the potential relevance of Löbner's (1985; 2011) four basic types of nouns for understanding definiteness/indefiniteness as associated with the [CL+N] construction in East and mainland Southeast Asian languages.

Even though the factors of semantics (animacy, uniqueness, relationality) and syntax (subject, object) clearly have an impact on the presence or absence of the classifier in contexts of definiteness and indefiniteness, we have evidence that discourse and information structure are stronger than these factors. The dominance of discourse is reflected in the very function of the classifier itself. As discussed in subsection 3.4.1, classifiers mark identifiability rather than uniqueness (cf. point (v.a), also cf. Li & Bisang 2012 on Sinitic). Thus, they express pragmatic definiteness rather than semantic definiteness in terms of Löbner (1985; 2011) or anaphoric ("strong") definiteness rather than unique ("weak") definite-

---

[22]In the case of Sinitic, Li & Bisang (2012) argue that the definiteness interpretation of subjects is due to a process of grammaticalization in which the definiteness properties of the topic position were passed on to the subject position (cf. the classical grammaticalization pathway from information structure to syntax in Givón 1979). In a similar way, the observation that postverbal [CL+N] constructions are preferably indefinite but do not exclude definiteness in Sinitic can be derived from the association of informational focus with the postverbal position (Xu 2004). As Lambrecht (1994: 262) points out, focus differs from topic inasmuch as it is not necessarily identifiable or pragmatically salient in discourse. For that reason, it is open to indefinite and definite interpretation even though the default interpretation is indefinite. If this analysis is true, one may argue that in Sinitic the classifier in [CL+N] is like a variable that takes on the [±definite] function that corresponds to its syntactic position if it is not overwritten by stronger factors. In Vietnamese, such a syntactic scenario turns out to be problematic because the classifier generally favours definite interpretation (cf. point (i)).

ness in terms of Schwarz (2009; 2013). In addition to the discourse-based definiteness expressed by the classifier, contrastive topics (subsection 3.4.2), as well as contrastive focus and focus particles (subsection 3.4.3), enhance the use of the [CL+N] construction. Thetic statements as another instantiation of information structure play an important role in the indefinite interpretation of [CL+N] in the subject position (subsection 4.1; also cf. (v.b)). Moreover, there are more specific discourse-based environments as mentioned in point (vi) which support the use of a classifier in contexts of indefinite interpretation (subsections 4.2 and 4.3). Finally, evidence of the dominance of discourse comes from data outside of our corpus. In order to disentangle the semantic effects of animacy vs. discourse effects associated with protagonists, we looked for narrative texts with inanimate protagonists. In the three texts we found, the inanimate protagonists generally occur in the [CL+N] construction (Quang forthcoming). One of the stories is about a flying carpet, which is already mentioned in the title, *Tấm thảm bay* [CL carpet fly] 'The Flying Carpet'.[23] After the protagonist is introduced by an indefinite construction of the type [one CL N], the noun *thảm* 'carpet' consistently occurs with a classifier. It is important to add in this context that the carpet has no anthropomorphic properties in the story, i. e., it does not act in any way. It is just the element that keeps the story going through many different events and episodes. Needless to say, such examples are hard to find in a corpus, no matter how large it is, because they are rare overall. The fact that even inanimate protagonists generally can take a classifier together with the findings summarized in (v) are good evidence for the dominance of discourse and information structure over semantics and syntax.

Taking these findings together, the classifier in [CL+N] is used as a variable whose use and interpretation depend on prominence in discourse and interact with factors from the morphosyntax-semantics interface. The details of that interaction will undoubtedly need more research. What is remarkable and makes the data on Vietnamese and other East and mainland Southeast Asian languages particularly relevant from a typological perspective is the observation that the different factors associated with (in)definiteness are well known, while crosslinguistic variation in how they interact is still under-researched. In Vietnamese, factors of discourse are particularly prominent. In order to further corroborate these observations and compare them with the situation in other mainland Southeast Asian languages, it is necessary to look at how classifiers are used in actual discourse in text corpora. We understand the corpus discussed here as a starting point for Vietnamese.

---

[23]The story was published by Viet Nam Education Publisher in 2003.

Acknowledgements

References

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

Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: Counting and beyond. In Jadranka Gvozdanovic (ed.), *Numeral types and changes worldwide* (Trends in Linguistics. Studies and Monographs (TiLSM), Vol. 118), 113–185. Berlin: Mouton de Gruyter.

Bisang, Walter & Yicheng Wu. 2017. Numeral classifiers in East Asia. *Linguistics* 55. 257–264.

Bornkessel-Schlesewsky, Ina & Matthias Schlesewsky. 2009. The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass* 3(1). 19–58.

Chafe, Wallace L. 1980. *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

Chen, Ping. 2004. Identifiability and definiteness in Chinese. *Linguistics* 42. 1129–1184.

Cheng, Lisa & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4). 509–542.

Croft, William. 2003. *Typology and universals.* 2nd edition. Cambridge: Cambridge University Press.

De Hoop, Helen A. & Andrej Malchukov. 2007. On fluid differential case marking: A bidirectional OT approach. *Lingua* 117. 1636–1656.

Dixon, Robert M. W. 1979. Ergativity. *Language* 55. 59–138.

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63. 805–855.

Emeneau, Murray B. 1951. *Studies in Vietnamese (Annamese) Grammar* (University of California Publications in Linguistics 8). Berkeley/Los Angeles: University of California Press.

Givón, Talmy. 1979. *On understanding grammar.* New York: Academic Press.

Greenberg, Joseph. 1972. Numerical classifiers and substantival number: Problems in the genesis of a linguistic type. *Working Papers on Language Universals* 9. 1–39. Department of Linguistics, Stanford University.

Jiang, L. Julie. 2015. Marking (in)definiteness in classifier languages. *Bulletin of Chinese Linguistics* 8. 319–343.

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.

König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective.* London/New York: Routledge.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55. 243–276.

Kuroda, Shichiro. 1972. The categorical and the thetic judgment: Evidence from Japanese syntax. *Foundations of Language* 9(2). 153–185.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents.* Cambridge: Cambridge University Press.

LaPolla, Randy J. 1995. Pragmatic relations and word order in Chinese. In Pamela A. Downing & Michael Noonan (eds.), *Word order in discourse*, 297–329. Amsterdam/Philadelphia: John Benjamins.

Li, Charles N. & Sandra A. Thompson. 1976. Subject and topic: A new typology of language. In Charles N. Li (ed.), *Subject and topic*, 457–489. New York: Academic Press.

Li, XuPing & Walter Bisang. 2012. Classifiers in Sinitic languages: From individuation to definiteness-marking. *Lingua* 122. 335–355.

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

Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.

Nguyen, Kim Than, Hai Thuy Ho & Duc Duong Nguyen. 2005. *Từ điển tiếng việt [Vietnamese Dictionary].* Ho Chi Minh City: Saigon Cultural Publisher.

Nguyen, Tuong H. 2004. *The structure of the Vietnamese noun phrase.* Boston University. (Doctoral dissertation).

Quang, Kim Ngoc. Forthcoming. *Vietnamese classifiers and (in)definiteness: A text-based analysis.* University of Mainz. (Doctoral dissertation).

Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25. 511–580.

Sasse, Hans-Jürgen. 1995. 'Theticity' and VS order: a case study. *Sprachtypologie und Universalienforschung* 48. 3–31.

Schwarz, Florian. 2009. *Two types of definites in natural language*. Amherst: University of Massachussets. (Doctoral dissertation).

Schwarz, Florian. 2013. Two types of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–558.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. New Jersey: Humanities Press.

Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asian languages. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 806–838. Oxford: Oxford University Press.

Simpson, Andrew. 2017. Bare classifier/noun alternations in the Jinyun (Wu) variety of Chinese and the encoding of definiteness. *Linguistics* 55(2). 305–331.

Sun, Chao-Fen & Talmy Givón. 1985. On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language* 61. 329–351.

Tran, Jennie. 2011. *The acquisition of Vietnamese classifiers*. Honolulu: University of Hawaii. (Doctoral dissertation).

Trinh, Tue. 2011. Nominal reference in two classifier languages. In Ingo Reich, Eva Horch & Dennis Pauly (eds.), *Sinn und Bedeutung 15: Proceedings of the 2010 annual conference of the Gesellschaft für Semantik*, 629–644. Saarbrücken: Saarland University Press.

Wang, Jian. 2015. Bare classifier phrases in Sinitic languages: A typological perspective. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 110–133. Oxford: Oxford University Press.

Wu, Yicheng & Adams Bodomo. 2009. Classifiers ≠ determiners. *Linguistic Inquiry* 40(3). 487–503.

Xu, Liejiong. 2004. Manifestations of informational focus. *Lingua* 114. 277–299.

Chapter 3

Preverbal (in)definites in Russian: An experimental study

Olga Borik

Universidad Nacional de Educación a Distancia (UNED), Madrid

Joan Borràs-Comes

Universitat Autònoma de Barcelona & Universitat Pompeu Fabra, Barcelona

Daria Seres

Universitat Autònoma de Barcelona

This paper presents an experimental investigation aimed at determining the exact nature of the relationship between type of interpretation (definite or indefinite) and linear position (pre- or postverbal) of bare nominal subjects of intransitive predicates in Russian. The results of our experiment confirm that preverbal position correlates with a definite interpretation, and postverbal position with an indefinite interpretation. However, we also discovered that the acceptance rate of preverbal indefinites is reasonably high. We suggest an explanation for the appearance of indefinites in preverbal subject position in terms of lexical accessibility, which is couched in general terms of D-linking.

1 Introduction

This paper is devoted to the study of bare singular nominals in Russian in pre- and postverbal subject position and a possible correlation between their (in)definiteness and their linear position in a sentence. Russian, as is well known, is a language without articles, i. e., a language that does not express definiteness as a grammatical category in a strict sense. This means that to establish the referential status of a bare nominal as a definite or an indefinite expression (a contrast

that seems to be perceivable for native speakers of Russian), the communication participants have to rely on a combination of clues and use various indicators provided both at a sentential and at a discourse level. In this paper, we are interested in establishing the role of a linear position of a nominal in this combination of factors that Russian uses to signal (in)definiteness.

To tackle this problem we conducted an experimental study, the empirical coverage of which is limited to subjects of stage-level intransitive verbs. In this study, native speakers of Russian were asked to judge the acceptability of sentences containing pre- and postverbal bare nominals in two types of contexts: definiteness- and indefiniteness-suggesting contexts. In definiteness-suggesting contexts we used anaphoric bare nominals, i. e., those that are linked to a referent in the previous context. This practical decision suggests a familiarity theory of definiteness (Christophersen 1939; Heim 1982) as a theoretical basis for the paper. The familiarity approach to definiteness is based on the idea that the referent of the definite description is known/familiar to both the speaker and the addressee. Definites are assumed to pick out an existing referent from the discourse, whereas indefinites introduce new referents (see specifically Heim 1982; Kamp 1981).

A different and very influential theory of definiteness is based on the uniqueness property of definite nominals (Russell 1905), which is usually taken to be part of the presupposition associated with a definite NP (Frege 1879; Strawson 1950). The two approaches are not mutually exclusive, and familiarity is sometimes claimed to be subsumed by uniqueness (see, for instance, Farkas 2002; Beaver & Coppock 2015). The basic idea behind the uniqueness approach is that a definite description is felicitous if, within a certain pragmatically determined domain, there is exactly one entity, in the case of singulars, or unique maximal set, in the case of plurals, satisfying the description.[1]

In this paper, we follow Farkas (2002), who introduced the notion of determined reference to "capture what is common to anaphoric and unique reference" (Farkas 2002: 221). Determined reference simply means that the value assigned to the variable introduced by a definite DP is fixed: there is no choice of entities that satisfy the descriptive content of a definite nominal. Although definite descriptions always have a determined reference, it can be achieved in different ways: definite DPs have determined reference if the descriptive content of a nom-

---

[1]Some other relevant theoretical notions related to definiteness in the current literature are determinacy (Coppock & Beaver 2015) and salience (von Heusinger 1997). We will not discuss these here, since a deep theoretical discussion of what it means to be definite is outside the scope of this paper. We limit our attention to one particular type of definite expression in this paper, although it is very well known that there are various types of definites (cf., for instance, Lyons 1999, for an overview).

inal (i.e., *cat* in 'the cat') denotes a singleton set relative to a context or if they are used anaphorically. In our experimental study, the nominal that appears in the definiteness-suggesting contexts is always anaphoric, by a link to a previous antecedent or by bridging.

The main conclusion drawn from our experiment is that linear word order in Russian cannot be considered the primary means for expressing definiteness and indefiniteness of bare nominals. Apart from the fact that we have confirmed a strong and clear correlation between linear position and interpretation of bare nominals, in the sense that preverbal bare subjects are mostly interpreted definitely and postverbal subjects indefinitely, we also report on another important result: some (not all) indefinite preverbal subjects are judged as acceptable by native speakers. It is this result that we are focusing on: in this paper, our aim is to ascertain what makes it possible or impossible to use a bare nominal subject in preverbal position in an indefiniteness-suggesting context. Thus, the main theoretical contribution of this paper consists in identifying requirements that facilitate the acceptability of what is considered an outcast, i. e., preverbal subjects with an indefinite interpretation. We propose that the general mechanism employed in licensing preverbal indefinite subjects is D-linking and identify some conditions for indefinites in Russian to be D-linked, justifying our proposal through an item-by-item analysis of all our preverbal contexts.

The rest of the paper is organised as follows. In section 2, we discuss the category of (in)definiteness and various means of expressing it, especially in those languages that lack articles and hence, do not have a straightforward way of signalling when a nominal is (in)definite. Our discussion is limited to Russian, a well-known representative of languages without articles. Section 3 is devoted to the experimental study that we have conducted with pre- and postverbal subjects of intransitive verbs. In this section, we outline the design and the methodology used in the experimental study, describe the results and present our interpretation of the results. In section 4, we discuss some theoretical issues that can be raised on the basis of the results of our experiment, and section 5 concludes the paper.

2 The category of (in)definiteness and its realizations

The category of definiteness (with two values, definite and indefinite) is mostly discussed in the literature in relation to articles, although it is often assumed that this category is, in fact, semantically universal and also present in those languages that do not possess formal means to express definiteness. The intuition is, indeed, that one of the differences in the interpretation of the nominal subject

in (1a) vs. (1b) in Russian corresponds to the contrast between (2a) and (2b) in English, where the (in)definiteness of the subject is overtly expressed. In Russian, even though the nominal subject appears in the same morphological form and without any additional markers in both sentences, the interpretation that the speakers are likely to attribute to the subject *koška* in (1a) by default seems to be indefinite, and thus comparable to the interpretation of *a cat* in (2a). However, the same nominal in (1b) is most likely to be interpreted as definite, and hence is comparable to the definite subject in the English example (2b).

(1) a. V uglu   spit koška.
        in corner.LOC sleeps cat.NOM

    b. Koška spit v uglu.
        cat.NOM sleeps in corner.LOC

(2) a. A cat is sleeping in the corner. / There is a cat sleeping in the corner.

    b. The cat is sleeping in the corner.

Thus, at least at first impression, definiteness forms part of the inventory of semantic contrasts/categories that can be expressed in Russian since the contrast between definite and indefinite readings of nominals can be easily perceived and understood by speakers. This observation is supported by the literature, where a common wisdom seems to be that languages without articles can express definiteness contrasts despite the absence of an article system. In fact, all the literature on definiteness in Russian simply assumes that it is entirely legitimate to talk about definite and indefinite readings. The only question that is discussed and debated is *how* (in)definiteness is expressed (see, for instance, Galkina-Fedoruk 1963; Pospelov 1970; Krylov 1984; Nesset 1999).

From a formal/compositional semantic perspective, a sentence like (3) needs some functional semantic operations to make sure that the result of combining a nominal phrase and a predicate in a simple intransitive sentence is well formed.

(3) Koška spit.
     cat.NOM sleep
     'A/the cat is sleeping.'

In formal semantics, common nouns like *cat* in English or *koška* in Russian are expressions of the type <e,t>, i. e., they denote a set of entities that can be characterized as cats.[2] Intransitive verbs are standardly given the same type

---

[2]See Chierchia (1998) for the claim that common nouns can be lexically of different logical types

<e,t>, as they denote a set of entities that sleep. Technically speaking, the two elements in (3) could be combined even though they are of the same type without the need to introduce any other semantic operations, for instance, by intersection, in which case we would end up with a set of entities that are sleeping cats. A way to combine the two elements without resorting to any type-shifting operations could be by pseudo-incorporation (e. g., Mithun 1984). However, in these cases the meaning predicted for the whole expression is far from the actual meaning of (3): (3) does not denote either a set of sleeping cats or a sleeping action as typically performed by cats. Moreover, the nominal itself does not exhibit any of the properties of pseudo-incorporated nominals (cf. Borik & Gehrke 2015 for an overview of such properties). The sentence in (3) is a typical predication, where something is said (asserted) about a cat entity. As a proposition, (3) can also be given a truth value. In order to properly derive the truth conditions of (3), we need to resort to type-shifting operations (Chierchia 1984; Partee 1987) which turn an argument (in this case, *cat*) into an entity <e> or a quantifier <<e,t>,t>.

In languages with articles, one of the functions that is attributed to an article or, in more general terms, a determiner, is shifting the noun denotation from a predicate type to an argument type. In particular, a type-shifting operation that the definite article 'performs' is called an *iota* shift and is formally defined as follows (see Heim 2011: 998):

(4) $[\![the]\!] = \lambda P{:}\exists x \forall y [P(y) \leftrightarrow x = y].\iota x.P(x)$
where ιx. abbreviates "the unique x such that"

It is reasonable to hypothesize that in Russian the same type-shifting rules can be applied as in English, although in the case of Russian the type-shifter itself is not lexically expressed. In fact, it has been proposed by Chierchia (1998) that exactly the same set of type-shifting operators that are used to formally derive argument types in languages like English can be employed in languages without articles to reflect various types of readings (entity type, predicate type or quantifier type) of nominal phrases. The proposal is quite attractive since it postulates a universal set of semantic operations that are used to model various denotations of nominal constituents. The only difference is that in some languages these operators are lexicalized (languages with articles), whereas in others they are not (languages without articles), as suggested, for instance, by Dayal (2004).

Thus, from a theoretical viewpoint, it is rather attractive to assume that a universal set of formal operators, called type-shifting operators, is postulated to

---

in different languages, although in his system English and Russian belong to the same group of languages, where the denotation of a common noun is taken to be of a predicate (i.e., <e,t>) type.

derive various readings of nominal phrases in different (possibly all) languages. Both definite and indefinite readings can then be derived by using appropriate type-shifting operations, regardless of language. It seems that we have ample empirical evidence from languages without articles like Russian that these readings do, indeed, exist, so that type-shifting operators are not vacuous, but give rise to various interpretations of nominal arguments, as illustrated in (1) and (3) above. However, in the absence of any obligatory lexical items that would reflect (in)definiteness of the corresponding nominal phrase, the question that arises is how we know when a nominal phrase is interpreted as a definite or as an indefinite one in a language like Russian.

2.1 Expressing (in)definiteness in Russian: lexical and grammatical means

Languages without articles possess various means to indicate the referential status of a nominal argument. In this section, we will review various means that can be employed in Russian to facilitate different (definite or indefinite) readings of a nominal.

First of all, there are lexical elements, including determiners, quantifiers and demonstrative pronouns,[3] that can be used to indicate whether the nominal they modify or combine with has a definite or an indefinite reading. Some examples of such lexical items are given in (5) below:

(5)   a.   Vasja znaet *ėtogo*   *studenta.*
           Vasja knows this.ACC.MASC student.ACC.MASC
           'Vasja knows this student.'

      b.   *Odna*   *znakomaja*   prixodila včera   v gosti.
           one.NOM.FEM acquaintance.NOM.FEM came   yesterday to guests
           'A (particular) friend came to visit yesterday.'

      c.   Vasju   iskala   *kakaja-to*   *studentka.*
           Vasja.ACC looked.for some.NOM.FEM student.NOM.FEM
           'Some student was looking for Vasja.'

      d.   Vasja opjat' kupil   *kakuju-nibud' erundu.*
           Vasja again bought some.ACC.FEM nonsense.ACC.FEM
           'Vasja bought some useless thing again.'

---

[3]Here we refer to a class of canonical, not pragmatic, demonstratives (cf. Elbourne 2008). Canonical demonstratives are strongly associated with definiteness in the literature (see, for instance, Lyons 1999; Wolter 2004; Elbourne 2008).

In (5a), the direct object *student* is preceded by a demonstrative, which gives the whole nominal phrase a definite interpretation: it is a particular, contextually unique and identifiable (possibly deictically) student that the nominal phrase refers to. In (5b), we are dealing with a specificity marker *odin* (lit. 'one', see Ionin 2013) and hence the whole noun phrase *one friend* is a specific indefinite. Similarly, the (postverbal) subject in (5c) is also a specific indefinite, although the marker here is different from the previous example. The last example, (5d), features a marker for non-specific low scope indefinites, so the object argument in this example is a weak indefinite.[4] In all these examples, there is a lexical determiner that indicates the definiteness status of a nominal argument, although these elements are really not like articles in the sense that it is never (or almost never) obligatory to use them.

Apart from lexical means, there are some grammatical tools in Russian that can affect the definiteness status of a nominal phrase. The two most well-known ones are case and aspect: both grammatical categories primarily affect the definiteness status of nominal arguments in direct object position. The influence of case-marking on referential properties can be demonstrated by the genitive/accusative case alternation on the direct object. For instance, mass nominal arguments of perfective verbs marked by the genitive case receive a partitive (indefinite) interpretation, whereas the same object in the accusative case can be interpreted as definite:

(6)   a. Vasja kupil   moloka.
            Vasja bought milk.GEN
            'Vasja bought (some) milk.'

    b. Vasja kupil   moloko.
            Vasja bought milk.ACC
            'Vasja bought (the) milk.'

Note, however, that the accusative case in (6b) allows for, but does not guarantee, a definite reading of the direct object *moloko* (milk.ACC), so that the observed effect is not strong enough to postulate a direct link between definiteness and case-marking.[5]

---

[4]Various indefiniteness markers in Russian are discussed in detail in the literature, especially in relation to specificity. See, for instance, Haspelmath (1997), Pereltsvaig (2000), Yanovich (2005), Geist (2008), Ionin (2013), etc.

[5]Speaking more generally, there is no correlation between case-marking and definiteness in Russian. There are languages that seem to exhibit such a correlation, especially with respect to direct object marking, such as Turkish, Persian (Comrie 1981/1989) and Sakha (Baker 2015).

Olga Borik, Joan Borràs-Comes & Daria Seres

As for aspect, the question of whether/how perfectivity influences the interpretation of a direct object in Slavic languages has been widely discussed in the literature (see Wierzbicka 1967; Krifka 1992; Schoorlemmer 1995; Verkuyl 1998; Filip 1999, etc.). It is often claimed (ibid.) that plural and mass objects of perfective verbs receive a definite interpretation,[6] whereas imperfective aspect does not impose any restrictions on the interpretation of a direct object. The claim is illustrated in (7) below:

(7) a. Vasja risoval   pejžaži.
         Vasja painted.ɪᴘꜰ landscapes.ᴀᴄᴄ
         'Vasja painted landscapes.'
     b. Vasja narisoval pejžaži.
         Vasja painted.ᴘꜰ landscapes.ᴀᴄᴄ
         'Vasja painted the landscapes.'

The effect of aspect on the interpretation of direct objects can be demonstrated very clearly in Bulgarian, another Slavic language, which, in contrast to Russian, does have a definite article. The example in (8) below, taken from Dimitrova-Vulchanova (2012: 944), illustrates that the definite article cannot be omitted if the verb is perfective:

(8) a. Ivan pi   vino.
         Ivan drank.ɪᴘꜰ wine.ᴀᴄᴄ
         'Ivan drank/was drinking wine.'
     b. Ivan izpi   vino*(-to).
         Ivan drank.ᴘꜰ wine.ᴀᴄᴄ-the
         'Ivan drank the wine.'

Thus, the correlation between the aspectual marking of a verb and the interpretation of its direct object seems, indeed, to be very strong. However, as illustrated above, perfective aspect is also compatible with an indefinite partitive interpretation if the object appears in the genitive case. Thus, in example (6a), the object is clearly indefinite and best translated as 'some (indefinite quantity of) milk' and not 'some of the milk'. Future, or non-past (Borik 2006) tense on a verb is another factor that can neutralize the effect of perfectivity: if the verb in (7b) is used in a non-past tense, the inferred definiteness of the direct object is considerably weakened or even invalidated. This means that the effect of aspect

---

[6]In the case of Verkuyl (1998), the terminology that is used is 'quantized', not 'definite'.

58

on definiteness of an internal argument is really just a tendency and might be overruled by other factors. But even in the strongest cases comparable to (7b), the correlation between perfectivity and definiteness in Russian is not absolute. Borik (2006: 92) provides an example where the internal argument of a perfective verb can have a non-maximal/existential interpretation:[7]

(9) Petja razdelil   ljudej   na dobryx i   zlyx.
     Petja divided.pf people.acc in kind.acc and mean.acc
     'Petja divided people into kind ones and mean ones.'

To summarize, we have seen that there are some grammatical factors, such as case or aspect, that can favour or facilitate a certain (definite or indefinite) interpretation of a nominal argument, but there are no strict correlations between definiteness and other grammatical categories. The lexical means that Russian possesses to signal (in)definiteness are only optional and cannot be semantically compared to articles. The interim conclusion is, then, that there is nothing so far in the grammatical system of Russian that would allow us to predict whether a nominal argument will necessarily be interpreted as a definite or an indefinite one.

Another factor often mentioned in the discussion of (in)definiteness in Russian is the effect of word order on the interpretation of nominal arguments, the phenomenon which underlies the experimental part of the paper. In the next subsection, we will briefly discuss word order in Russian and its (potential) relation to definiteness, and provide motivation for the experiment that will be reported in section 3 of the paper.

2.2 The effects of word order on the interpretation of nominal arguments

Russian is a classic example of a so-called 'free word order' language, i. e., a language where the linear order of the elements in a sentence is determined not so much by grammatical functions like subject and object, or grammatical properties like case assignment, but by the requirements imposed by discourse and information structure (see Mathesius 1964; Sgall 1972; Hajičová 1974; Isačenko 1976; Yokoyama 1986; Comrie 1989, among others). However, more cautious typological sources always point out that the 'free' word order is to a large extent an illusion, since various permutations of sentence constituents are usually not entirely

---

[7]In fact, Borik (2006) claims that the interpretation in this case is 'generic'. However, since the sentence itself is not interpreted generically but rather refers to an episodic event, 'existential interpretation' is a more accurate term. We thank one of the reviewers for pointing this out to us.

Olga Borik, Joan Borràs-Comes & Daria Seres

free but guided by some pragmatic or information structure principles (see, for instance, Dryer 2013). For these languages, the connection is often made between the linear position of a nominal argument and its definiteness status. In particular, it is often stated in the literature that preverbal (subject) position is strongly associated with a definite interpretation (Pospelov 1970; Fursenko 1970; Krámský 1972; Chvany 1973; Szwedek 1974; Topolinjska 2009, etc.), whereas nominals in postverbal position are likely to be interpreted as indefinites. This descriptive generalization is primarily assumed to hold for subjects, as the canonical word order in Russian is SVO, and objects, unless they are topicalized, follow the verb rather than preceding it.

The relationship between definiteness and preverbal subjects is often mediated by topicality, a notion that plays a crucial role in the interpretation of arguments in languages with a (relatively) free word order. Both preverbal subjects and objects are considered topics when they appear in sentence-initial position (Jasinskaja 2014). As illustrated in the examples below, both subject (see (10)) and object (see (11)) in the leftmost position can also be left-dislocated (creating, arguably, a bi-clausal structure), a construction that we consider to be a reasonable, although not clear-cut diagnostic for topichood (Reinhart 1981).

(10)   a.   Tolya včera   razgovarival s   Anej.
             Tolya yesterday talked.IPF   with Anya
             'Tolya yesterday talked to Anya.'

      b.   Čto kasaetsja Toli, to on včera   razgovarival s   Anej.
             what concerns Tolya that he yesterday talked.IPF   with Anya
             'As for Tolya, he talked to Anya yesterday.'

(11)   a.   Varenje ja včera   s'el.
             jam   I   yesterday ate.PF
             'I ate the jam yesterday.'

      b.   Čto   kasaetsja varenja, to   ja ego včera   s'el.
             what concerns jam   that I   it   yesterday ate.PF
             'As for the jam, I ate it yesterday.'

The type of topic illustrated in (10a) and (11a) is called aboutness topic (Reinhart 1981) or, what we believe to be essentially the same phenomenon, internal topics (King 1995). The connection between definiteness and topicality is based on a descriptive generalization that is accepted in a lot of semantic literature on topics in general, namely, that elements that appear in topic position can only be referential, i. e., definite or specific indefinite (see Reinhart 1981;

Erteschik-Shir 1997; Portner & Yabushita 2001; Endriss 2009, etc.). An appealing intuitive idea behind this generalization is that if there is no entity that the nominal topic refers to, this expression cannot be aboutness topic because then there is no entity to be talked about.

Nevertheless, a number of examples from various Romance languages have been brought up in the literature to show that a topicalized left-dislocated element can, in fact, be interpreted non-specifically. The following examples from Leonetti (2010) illustrate the phenomenon in Spanish (12a) and Italian (12b):

(12) a. Buenos vinos, (los)   hay   en Castilla.
          good   wines (CL.MASC.PL) have.PRS.3SG in Castile
          'There are good wines in Castile.'

      b. Libri   in inglese, (li/ne)   può   trovare al
          books in English (CL.MASC.PL/CL.PART) can.PRS.3SG find   on.the
          secondo piano.
          second   floor
          'English books can be found on the second floor.'

As suggested by Leonetti (2010), non-specific or weak indefinites are highly restricted in topic position. He identifies two conditions that must be met to allow for non-specific indefinites to appear as topics. First, they can be licensed by certain kinds of contrast that cannot lead to a specific reading. This condition has to do with intonation and stress, factors that fall outside the scope of the discussion in this paper. Second, they can be licensed in the sentential context with which the topic is linked. In other words, this second condition means that what matters for licensing non-specific indefinite topics is the presence of supporting context. In general, the examples in (12) illustrate that the correlation between topic and definiteness and/or topic and referentiality is not a strict dependency but rather a strong tendency.

For Russian, as well as for other languages with free word order, it is important to dissociate the effects that can be attributed to topicality from those that can (potentially) arise merely from word order. In particular, the question that we address in this paper is whether the linear position of a subject, regardless of topichood, correlates with its definiteness or not. Therefore, our experimental items include preverbal subjects which are not topics, i.e., which do not appear in a sentence-initial position. It has been claimed in the literature that preverbal subjects that are not topics, for instance, preverbal subjects of thetic sentences, can be both definite and indefinite (cf. Geist 2010, among others), but

the results of our experiment suggest that there is nonetheless a strong dependency between linear position and interpretation. In particular, we will show that indefinites have a relatively low acceptance rate when they appear preverbally in non-topical contexts. Just like (weak) indefinite topics, preverbal non-topical indefinites seem to still need contextual support, so the conditions for licensing indefinites in preverbal position appear to be quite rigorous. Thus, the generalization seems to be that preverbal indefinites need special contextual conditions to facilitate their use, regardless of whether they are topics.

3 The experimental study

The relationship between the syntactic position of a bare nominal and its interpretation has been found in other languages (e. g., Cheng & Sybesma 2014, for Mandarin Chinese); it has even been claimed that the pattern where the preverbal nominal is interpreted definitely and the postverbal nominal is interpreted indefinitely is universal (Leiss 2007). However, there have not been many experimental studies based on articleless languages to ascertain how speakers interpret bare nominal subjects in preverbal and postverbal position. Some of the most relevant experimental studies that have been conducted for Slavic languages are discussed in the next subsection. The scarcity of experimental work concerning the interpretation of bare nominals in Slavic languages in general and in Russian in particular motivated our study of Russian bare plural subjects.

3.1 Previous experiments

The recent experimental studies on Slavic languages that we are aware of are the study of bare singular NPs in Czech by Šimík (2014), a statistical analysis based on Polish and English texts by Czardybon et al. (2014) and Šimík & Burianová (2018), who conducted a corpus study of bare nominals found in pre- and postverbal position in Czech. All the studies, even though methodologically different, show that there is a quite strong correlation between word order and the interpretation of nominal arguments.

     Šimík's (2014) experiment tested the preference for a definite or an indefinite reading of an NP in initial or final position in a sentence. The study demonstrated that the initial position (topicality) of the subject increased the probability of a definite interpretation; however, it was not a sufficient force to ensure this type of reading. Even though the indefinite interpretations were selected less for

NPs in initial position than in final position, they were still not excluded. Moreover, indefinite interpretations were overall preferred over definite ones.

A comparative study of Polish translations of English original texts by Czardybon et al. (2014) aimed to provide a quantitative assessment of the interaction between word order and (in)definiteness in Polish. The results of this quantitative evaluation support previous theories about the correlation between the verb-relative position and the interpretation of bare nominals: preverbal position is strongly associated with definiteness and postverbal position is connected to the indefinite reading of an NP. Nevertheless, the study revealed quite a high number of preverbal indefinite NPs, which the authors were not expecting (Czardybon et al. 2014: 147-148). However, as pointed out by Šimík & Burianová (2018), Czardybon et al. (2014) did not distinguish between preverbal and sentence-initial position, which complicates the interpretation of their results considerably.

Some important and relevant findings concerning the relationship between definiteness of a nominal argument and its linear position in a sentence are reported in Šimík & Burianová (2018), who conducted a corpus study and discovered that in Czech, clause-initial position shows very high intolerance towards indefinite nominal phrases. Šimík & Burianová (2018) argue that definiteness of bare nominals in Slavic is affected by an absolute (i. e., clause-initial vs. clause-final) but not a relative (i. e., preverbal vs. postverbal) position of this nominal in a clause. Our experimental findings, which will be described in the next section, seem to contradict this conclusion. In particular, we find that preverbal indefinites in non-initial position have much lower acceptability than postverbal ones. We therefore argue that preverbal indefinites need additional anchoring mechanisms to be activated, which would ensure their successful use in a given context. We will propose that this anchoring mechanism is D-linking, a general discourse coherence principle that can be defined by a set of specific conditions.

All the studies reviewed in this section demonstrate that, at least to some extent, NPs with an indefinite interpretation do appear preverbally, where they are not generally expected. Our experiment will also confirm this result.

3.2 Overall characteristics of the experiment

This section provides a general description of the experimental study we conducted. As mentioned above, the aim of our study was to investigate the relationship between the interpretation of bare nominals in Russian and their position in the sentence (preverbal or postverbal), which relies on the long-standing assumption that word order in articleless Slavic languages is one of the means of expressing (in)definiteness. The main goal of the experimental study was to see whether the

claim that preverbal bare subjects are interpreted definitely, while postverbal bare subjects are interpreted indefinitely correlates with native speaker judgements.

Given that we limited our study to anaphoric definiteness, our initial hypothesis can be formulated as follows:

(13)   *The preverbal position of the bare subject expresses definiteness (familiarity) and the postverbal position expresses indefiniteness (novelty).*

In order to verify this initial hypothesis, a survey was created. The interpretation of bare subject NPs was examined using an Acceptability Judgement Test (AJT) with a scale from 1 (not acceptable) to 4 (fully acceptable). The survey was administered online using the SurveyMonkey software. The items were presented to participants visually and acoustically, so as to avoid a possible change in the interpretation due to intonation, as it has been claimed in the literature that if a preverbal noun receives a nuclear accent, it may be interpreted indefinitely (Pospelov 1970; Jasinskaja 2014, among others). Potentially, the effect of intonation may be stronger than the word order restriction described above (see the initial hypothesis). Thus, we considered it important to exclude variation in judgement caused by intonation and all the experimental items were recorded with the usual, most neutral intonation contour used for statements in Russian (intonation contour 1, cf. Bryzgunova 1977). This intonation contour is characterized by a flat, level pitch before the stressed syllable of the intonational nucleus, i. e., the stressed syllable of the most informative word in a sentence. In our examples, the stress was always on the last word of the sentence.

A total of 120 anonymous participants took part in the survey. Demographic information about the participants was collected in a pre-survey sociological questionnaire. All participants claimed to be native Russian speakers; the gender distribution was 102 women, 17 men, one non-binary; the mean age (in years) was 36.59 (SD = 8.55); 91 participants claimed to have received a university degree in language-related fields.

The experimental items contained a bare subject nominal in a preverbal or postverbal position. All predicates were stage-level, according to Carlson's (1977) classification, expressed by an intransitive verb. All subject nominals were plural for the sake of uniformity; however, we expected the effects found in the course of the experiment to be manifested in the case of singular nominals as well. The experimental sentences were presented in a brief situational context, which suggested either novelty (associated with indefiniteness) or familiarity (associated with definiteness) of the subject. A total of 48 items were presented to participants: 16 definiteness-suggesting (8 preverbal and 8 postverbal) plus 16

indefiniteness-suggesting (8 preverbal and 8 postverbal) experimental scenarios, and 16 fillers. The average answer time was 22 minutes.

Below we provide some examples of our experimental items.

(14) *Preverbal subject in an indefiniteness-suggesting context:*

Ėto že pustynja, ėto samaja nastojaščaja pustynja. V ėtoj mestnosti do six por ne bylo ni odnogo živogo suščestva. No na prošloj nedele **pticy** *prileteli*.[8]
'It's a desert, it's a real desert. In this area there has never been a living creature. But last week birds came (lit. **birds** *came.flying*).'

(15) *Postverbal subject in an indefiniteness-suggesting context:*

Čto-to strannoe stalo proisxodit' v našej kvartire. V kuxne vsegda bylo očen' čisto, nikogda ne bylo ni odnogo nasekomogo. No nedelju nazad *obnaružilis'* **tarakany**.
'Something strange started happening in our flat. It has always been very clean in the kitchen, there has never been a single insect. But a week ago cockroaches were found (lit. *found.themselves* **cockroaches**).'

(16) *Preverbal subject in a definiteness-suggesting context:*

Kogda Katja i Boris vernulis' iz otpuska, oni obnaružili, čto ix dom ograblen. Pervym delom Katja brosilas' v spal'nju i proverila seif. Ona uspokoilas'. **Dragocennosti** *ležali* na meste.
'When Katja and Boris came back from holiday, they discovered that their house had been burgled. First of all, Katja rushed into the bedroom and checked the safe. She calmed down. The jewellery was still there (lit. **jewelleries** *lay* in place).'

(17) *Postverbal subject in a definiteness-suggesting context:*

Oživlenije spalo, publika potixon'ku potjanulas' domoj. "Počemu vse uxodjat?" – sprosil Miša. "Gonki zakončilis'. V garaži *vernulis'* **mašiny**."
'The agitation declined, the public slowly started going home. "Why is everybody leaving?" Misha asked. "The race has finished. The cars have returned to their garages." ' (lit. to garages *returned* **cars**).

In the following section we discuss the results of the experiment.

---

[8]In order to make the examples easier to understand, the bare NP subject is in bold, while the verb is in italics. This marking does not reflect any stress pattern.

Olga Borik, Joan Borràs-Comes & Daria Seres

3.3 General results

A total of 3,840 data points were collected (120 participants × 2 definiteness conditions [indefinite, definite] × 2 positions in which the NP appeared in the sentence with respect to the verb [preverbal, postverbal] × 8 scenarios). These responses were analyzed using a Linear Mixed Model using the GLMM interface from IBM SPSS Statistics 24.

The Linear Mixed Model was applied to the data. The model was defined with Participant as the subject structure and Situation × Position as the repeated measures structure (Covariance Type: Diagonal). The participants' perceived acceptability of the sentences was set as the dependent variable. The fixed factors were Definiteness, Position, and their interaction. Regarding the random factors, a random intercept was set for Participant, with a random slope over Position (Covariance Structure: Variance Components).

The two main effects were found to be significant: Definiteness, $F(1, 3829)$ = 44.700, $p < .001$, such that indefinite sentences obtained significantly more acceptability than definite sentences (diff = .164, SE = .024, $p < .001$), and Position, $F(1, 3829)$ = 14.236, $p < .001$, indicating that preverbal NPs obtained more acceptability than postverbal NPs (diff = .113, SE = .030, $p < .001$).

The interaction Definiteness × Position was found to be significant, $F(1, 3829)$ = 4958.853, $p < .001$, which could be interpreted in the following two ways. On the one hand, in preverbal position definites were more adequate than indefinites (diff = −1.561, SE = .035, $p < .001$), and in postverbal position indefinites were more adequate than definites (diff = 1.888, SE = .034, $p < .001$). On the other hand, indefinites were found to be more adequate in postverbal rather than in preverbal position (diff = −1.612, SE = .037, $p < .001$), while definites were found to be more adequate in preverbal rather than in postverbal position (diff = 1.837, SE = .040, $p < .001$). Figure 1 shows the mean perceived acceptability that the participants ascribed to the experimental items on the 4-point Likert scale, from 1 (not acceptable) to 4 (fully acceptable).
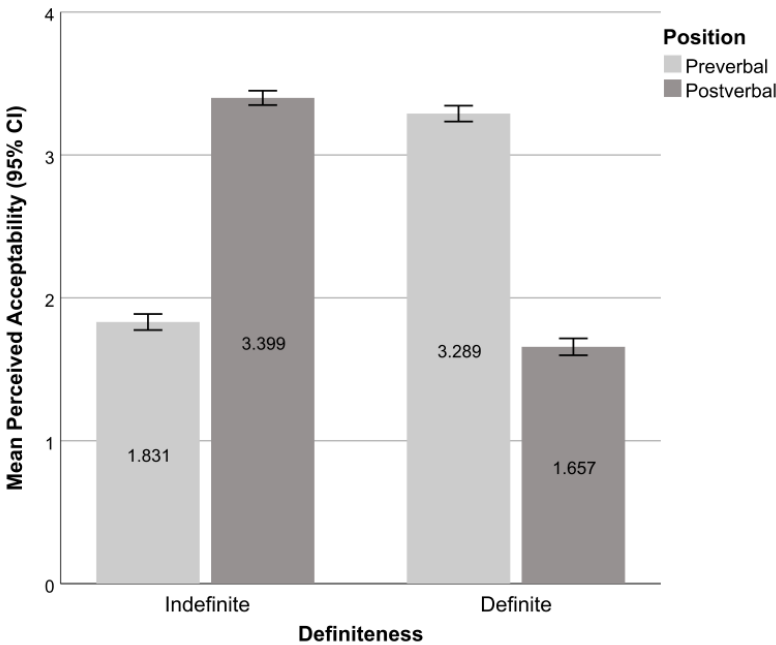
Figure 1: Average perceived acceptability that our participants attributed to the experimental sentences. Error bars depict the 95% confidence interval.

The most perceptible result seen from the graph is that the participants favoured two out of the four possible combinations of Definiteness and Position, i. e., postverbal subjects in indefiniteness-suggesting contexts (M = 3.399, SD = .791) and preverbal subjects in definiteness-suggesting contexts (M = 3.289, SD = .874), giving substantially lower ratings to preverbal subjects in indefiniteness-suggesting contexts (M = 1.831, SD = .885) and postverbal subjects in definiteness-suggesting contexts (M = 1.657, SD = .932).

Besides the optimal combinations (preverbal NP + definiteness-suggesting context and postverbal NP + indefiniteness-suggesting context), additional statistically significant results were obtained. Firstly, an overall superior acceptability for NPs in indefiniteness-suggesting contexts (regardless of the syntactic position of the NP) as compared to definiteness-suggesting ones was observed. Secondly, the acceptability of bare nominals in preverbal position was higher compared to the postverbal position, regardless of type of preceding context.

The results, in our view, can be interpreted in the following way. First of all, there is quite a strong preference for interpreting preverbal NPs definitely and

postverbal NPs indefinitely. However, there is no clear one-to-one correspondence, which suggests that the linear position of a subject nominal in Russian cannot be considered a means of expressing its definiteness/indefiniteness. So, our initial hypothesis has to be modified. Instead of saying that the word order encodes the referential status of a nominal (i. e., its definiteness or indefiniteness), we think the results only show that preverbal nominal subjects are *much more likely* to be interpreted as definites. Indefinites are not rare in this position either and their acceptability is fairly high, so our next question is what the factors are that influence speakers' judgements in the case of preverbal indefinites.

In an attempt to answer this question we looked at our preverbal definiteness - and indefiniteness-suggesting contexts one by one and tried to analyze every preverbal context that we had used in our experimental study. The results that we obtained are reported in the next section.

3.4 Item-by-item analysis of preverbal subjects

One of the main theoretical questions that we try to answer in this paper is what makes it possible for a particular nominal to appear in a preverbal subject position. In search for a possible answer, we looked at the information status of the subject NPs in the experimental sentences. Baumann & Riester (2012) claim that, for an adequate analysis of the information status of a nominal expression occurring in natural discourse, it is important to investigate two levels of givenness: referential and lexical. The authors propose a two-level annotation scheme for the analysis of an NP's information status: *the RefLex* scheme. In this article we adopt this scheme in order to investigate the correlation between acceptability of an item in preverbal position and its information status.

3.4.1 Definiteness-suggesting contexts

In definiteness-suggesting contexts, the subject NPs can be labelled, according to Baumann & Riester's RefLex scheme (2012: 14), as *r-given* or *r-bridging* at a referential level. The *r-given* label is used when the anaphor co-refers with the antecedent in the previous discourse. *R-bridging* is assigned when the anaphor does not co-refer with an antecedent but rather depends on the previously introduced scenario. At a lexical level, the items can be classified (Baumann & Riester 2012: 18-19) as *l-given-syn* (the nouns are at the same hierarchical level, i. e., synonyms), *l-given-super* (the noun is lexically superordinate to the nominal antecedent), *l-accessible-sub* (the noun is lexically subordinate to the nominal antecedent) or

l-accessible-other (two related nouns, whose hierarchical lexical relation cannot be clearly determined).

Table 1 represents the experimental scenarios with definiteness-suggesting contexts. It provides the anchor nominal from the previous context, the target nominal (the preverbal subject NP from the experimental sentence), the RefLex labels of the target nominal, the mean acceptability given (M; in a 0 to 1 scale)[9] and the standard deviation (SD) acceptability figures for each item.

Table 1: Annotation of target nominals in definiteness-suggesting contexts

	Previous context	Target nominal	RefLex annotation	M	SD
1	boy and girl[a]	children	r-given, l-given-syn	.8333	.2520
2	family of tigers	animals	r-given, l-given-super	.7750	.2806
3	safe	jewellery	r-bridging, l-accessible other	.8833	.2102
4	canary and parrot	birds	r-given, l-given-super	.8418	.2166
5	crucians	fishes	r-given, l-given-super	.6863	.3016
6	family silverware	cutlery	r-given, l-given-syn	.7583	.2930
7	Plato and Aristotle	philosophers	r-given, l-given-super	.5972	.3372
8	races	cars	r-bridging, l-accessible-sub	.7306	.3155

[a]We are not using articles here as, naturally, they are absent in the Russian examples.

As can be seen from Table 1, the acceptability of preverbal nominals in definiteness-suggesting contexts is quite high and fairly uniform. This is an expected result as preverbal position is strongly related with familiarity/identifiability of the referent and the degree of givenness, which is high in all cases (as can be seen from the labels). So, it is natural for NPs to appear preverbally in definiteness-suggesting contexts, when they are anaphorically or situationally related to an antecedent in a previous context.

The item with the lowest (although still high, in absolute terms) acceptability is 7, given in (18):

[9]The original acceptability variable was changed from 1-4 to 0-1 for reasons of clarity. This change was the result of the following formula: (acceptability – 1)/3. We consider it to be easier to interpret what an acceptability score of .4 on a 0-1 scale represents than the equivalent score of 2.2 on a 1-4 scale, which might be misconceived as if it was a 0-4 scale, thus indicating more than a half of accepted readings.

(18)   Sredi mnogočislennyx antičnyx prosvetitelej, otmetivšixsja v istorii, možno vydelit' neskol'ko naibolee važnyx. Platon i Aristotel' izvestny vo vsëm mire. **Filosofy** *žili* v Drevnej Grecii.

'Among numerous classical thinkers that left their trace in the history it is possible to distinguish a few most important ones. Plato and Aristotle are known all over the world. The philosophers lived in Ancient Greece (lit. **philosophers** *lived* in Ancient Greece).'

In terms of its information status, the bare nominal subject *philosophers* is not really different from the subjects of other items: it is *r-given*. A lower acceptability rate must then be due to other factors, e. g., the use of proper names or attributing a generic type of interpretation to the last sentence (i. e., 'In general, philosophers lived...'), which would cancel the anaphoric connection.[10]

Thus, apart from one item discussed above (item 7), all the definiteness-suggesting contexts show the same result: a high acceptability rate for the preverbal bare nominal subject.

3.4.2 Indefiniteness-suggesting contexts

In all indefiniteness-suggesting contexts, the existence of referents was negated; thus, the novelty of the target nominal was presupposed. Using Baumann & Riester's (2012: 14) annotation scheme, at a referential level all the target NPs are classified as *r-new*, i. e., specific or existential indefinites introducing a new referent. At a lexical level, they are either *l-accessible-sub* or *l-accessible-other*. Table 2 presents the annotation results for bare nominals in preverbal position in indefiniteness-suggesting contexts.

---

[10]It is interesting to note that a similar effect has been observed for an item with a postverbal subject in a definiteness-suggesting context. While other items with definite postverbal subjects were given low acceptability as expected, the acceptability of this particular item was quite high (M = .4667, SD = .3441). The English translation of the item is given in (i):

(i)   I love birds and I advise all my friends to have at least one feathered pet. They are generally undemanding, although sometimes they make noises and give you extra trouble. At home I have *lit.* **canary and parrot**. Yesterday I forgot to close the cage's door, and all day long *lit.* around room *flew* **birds**.

Our hypothesis is that the informants processed the antecedent and the anaphor NPs in this example as non-co-referential, therefore interpreting the target NP as referentially new, which made it possible for the subject to be accepted in postverbal position.

Table 2: Annotation of target nominals in indefiniteness-suggesting contexts

	Previous context	Target nominal	RefLex annotation	M	SD
1	no rodents	mice	r-new, l-accessible-sub	.3833	.3195
2	no insects	cockroaches	r-new, l-accessible-sub	.2167	.2755
3	empty street	people	r-new, l-accessible-other	.1750	.2766
4	no fruit	bananas	r-new, l-accessible-sub	.2778	.2778
5	no living creatures	birds	r-new, l-accessible-sub	.2861	.2940
6	no mail	postcards	r-new, l-accessible-sub	.3056	.3163
7	no domestic animals	cats	r-new, l-accessible-sub	.3194	.2847
8	no wild animals	wild boar	r-new, l-accessible-sub	.2521	.2709

As can be seen from Table 2, the acceptability of preverbal NPs in indefiniteness-suggesting contexts is uniformly low, but high enough to be statistically significant (see section 3.2). All these NPs are referentially new. However, it should be pointed out that at a lexical level, the target nominals in indefiniteness-suggesting contexts are accessible, being a subset of a superset mentioned in the previous context.

The item that received the lowest ranking in Table 2 is item 3 (M = .1750, SD= .2766), which has a slightly different information status label at a lexical level. It has an *l-accessible-other* label, which means that, unlike other items with a clear lexical relation of hyponymy, the hierarchical relation between the context and the target NP cannot be clearly established in the given scenario. Item 3 is provided in (19):

(19)   Bystro stemnelo, nastupil večer. Na ulice bylo tixo i pustynno. Vdrug iz-za ugla **ljudi** vyšli.
       'It got darker, the night came very quickly. *Lit.* In the street it was silent and empty. Suddenly from around the corner *lit.* **people** came out.'

As opposed to other experimental scenarios, in this context there is no NP to which the target nominal *people* could be anchored. Even though it can be linked to the whole context, given our common knowledge that people usually walk in the streets, this vague type of contextual support does not seem to be enough to 'license'[11] the bare nominal *people* to appear in preverbal position. Even though it is just one example, we believe that the lower acceptability rate

---

[11]We use the term 'license' here in a loose sense, without appealing to anything as strict as 'licensing conditions', the way they are understood in syntax.

of this example might not be accidental. In the next section we will discuss the factors that could make this sentence different from the other experimental items in the same group.

4 Some theoretical considerations

We begin this final section of the paper by suggesting a tentative answer to the question posed in the previous sections: what are the conditions that bare indefinites have to meet to be accepted in preverbal position? Having analyzed the data in all preverbal contexts, we can propose a plausible hypothesis as an answer to this question, although the validity of this hypothesis should be further confirmed in future empirical and experimental studies.

An item-by-item analysis of our experimental scenarios suggests that if an item is *r-given*, it has a tendency to appear preverbally, and this combination (i. e., *r*-givenness and preverbal position) is judged highly acceptable by native speakers of Russian. This is illustrated by the item-by-item analysis of our definiteness-suggesting contexts. If, however, a nominal is *r-new*, it is judged much less acceptable in preverbal position, even though it is still tolerable: the acceptability rate for these items was about 1.8 on a 4-point scale, as we saw in section 3.3, where the general results of the experiments were discussed. What our data seems to indicate is that it is not only referential givenness but also accessibility at a lexical level that plays a significant role in licensing bare nominals in preverbal position. Thus, if a bare nominal is *r-new*, it can still appear preverbally in those cases where it establishes a clear lexical connection with a nominal phrase in the previous context. However, in the example where the connection between the previous context and the target item is looser and the item can only be classified as *l-accessible-other* (i. e., a target nominal can only be pragmatically related to the whole context), the acceptability rate drops and the item is judged close to unacceptable.

It might be too early to draw any far-reaching theoretical conclusions on the basis of just one experiment with 16 test items. However, we believe that the results we obtained in our experimental study for preverbal bare nominals in Russian at least allow us to identify some conditions that seem to facilitate the use of bare nominal phrases in indefiniteness contexts in preverbal position in Russian: *r*-givenness and *l*-accessibility. We would like to suggest that these conditions could be connected with a much broader phenomenon, which might serve as a general explanation for a reduced and restricted, but still accepted, appearance of indefinite nominal phrases in preverbal position. The phenomenon that we refer to is called D-linking.

Pesetsky (1987) described discourse linking (or D-linking) as a phenomenon where one constituent is anchored to another one in the preceding discourse or extralinguistic context. Dyakonova (2009: 73), building on this idea, gives the following definition of D-linking:

(20) A constituent is D-linked if it has been explicitly mentioned in the previous discourse, is situationally given by being physically present at the moment of communication, or can be easily inferred from the context by being in the set relation with some other entity or event figuring in the preceding discourse.

As can be seen from this definition, D-linking is a rather broad phenomenon that allows for various connections to be established between a constituent X and the preceding discourse or a situational context. We suggest that this general phenomenon could be split into a set of specific conditions that would allow us to achieve a more precise characterisation of D-linking.

As was pointed out in section 2.2, discourse support seems to play a role in licensing non-specific (weak) indefinite nominals in topic position in Romance languages. For instance, Leonetti (2010) identifies two conditions under which non-specific indefinites appear as topics in Romance languages: contrast and contextual support. We suspect that the latter could fit into what we describe as D-linking although the precise characterization of what it means to be contextually supported is yet to be established.

Our experiment has shown that native speakers of Russian, a language which does not encode (in)definiteness by any grammatical means, can accept an indefinite interpretation of a bare nominal in preverbal position, even though a general acceptability rate for preverbal indefinites is much lower than for preverbal definites. What we have suggested in this paper is that for indefiniteness contexts, not only referential, but also lexical linking to a previous nominal element can play a role. Those nominals that were strongly supported by the previous contexts by lexical relations such as hyponymy/hyperonymy are judged more acceptable than those which do not have this type of support. This, of course, is almost directly captured by the definition of D-linking given in (20): in all our test sentences, the preverbal nominals in indefiniteness-suggesting contexts were lexically accessible, as they were in the set relation with an entity from the preceding context, except for item 3 (which obtained the lowest acceptability). This fact indicates that it would be plausible to explore the role of D-linking principle(s) for a general account of the distribution of bare nominals with indefinite readings in Russian.

Olga Borik, Joan Borràs-Comes & Daria Seres

To conclude this section, we would like to raise another theoretical issue that has come to our attention, both while conducting the experiment that we have reported on here and in the course of the interpretation of the results. It concerns the notion of definiteness and the status of our test items with respect to this category.

As we pointed out in the introduction, the debate on what definiteness actually means in semantic terms continues, although here we follow Farkas (2002) in assuming that the familiarity and the uniqueness approaches to definiteness could converge. In the experiment reported on here, in the definiteness-suggesting contexts, we test nominal phrases that are anaphorically linked to a referent in the preceding discourse, and we consider our experimental items to be fully legitimate candidates for definite nominals in the definiteness contexts because they are familiar to the speaker. However, all our anaphoric nominals in definiteness-suggesting contexts are also given, so the question that presents itself is whether the results of the experiment are influenced by the givenness status of the tested items.[12]

Givenness is a category related, although not equivalent, to definiteness. An element is given if there is an antecedent for it in the preceding discourse, so givenness is an information-structural category that is also closely related to anaphoricity. Any constituent of a sentence can have the status of 'given', including, of course, nominal arguments.

The relationship between definiteness and givenness is not straightforward: in principle, both definite and indefinite arguments can be either given or new. For instance, any contextually unique definite mentioned for the first time is not given but new (e. g., *The UV is very high today, The head of the department just called me*), whereas any anaphoric definite is given. Crucially, however, the given/new status seem to correlate with stress: deaccentuation and word order are common ways to indicate givenness of a certain constituent (cf. Krifka 2008).[13] As for Slavic languages, Šimík & Wierzba (2015) present a thorough study of the interaction between givenness, word order and stress in Czech.

As we have already mentioned, in our experiment we tried to eliminate the stress factor, by recording all our example sentences with a neutral intonation, flat pitch, and a phrasal stress at the end of a sentence. Hence, all preverbal items (in definiteness and indefiniteness contexts) were unstressed and postverbal items

---

[12]We thank a reviewer for bringing up this question and for suggesting that we look into the role of givenness in the distribution of nominal arguments.

[13]Although see Rochemont (2016) for the claim that only salient-based givenness is associated with deaccenting.

(in definiteness and indefiniteness contexts) were stressed only when they also appeared in a sentence-final position. It might be that givenness is the factor that influenced the acceptability judgements in our experiment, especially in the case of nominals in definiteness contexts, because the speakers might have been less willing to accept a postverbal stressed given nominal, since the nominal appeared in sentence-final position. However, to properly answer this question we need to design an experiment with postverbal definites that appear in sentence-final and non-final position, and also manipulate stress. Stress might be particularly important for indefinite nominals, as it has been noted in the literature that stressed indefinites become more acceptable in preverbal position. All in all, we think that studying the role of givenness versus definiteness in the distribution of bare nominal arguments is an exciting task for a (near) future project.

5 Conclusions

In this paper, we have discussed the relationship between the definiteness status of a bare nominal and its linear position in a sentence in Russian. We have confirmed that, according to the results of the experiment that we conducted with native speakers of Russian, the general tendency is, indeed, to associate preverbal position with a definite interpretation and postverbal with an indefinite one, although it cannot be stated that this connection is a strict correspondence. Consequently, we cannot say that linear position 'encodes' definiteness or indefiniteness: the observed correlations are tendencies rather than strict rules.

One of the other results of our experiment is the reasonably high ranking that is assigned to bare nominals with an indefinite interpretation that appear in preverbal position. We carried out an item-by-item analysis of all the preverbal nominals with the aim of identifying a specific condition or a set of specific conditions that would make indefinite nouns acceptable in this position. Our conclusion was that the condition has to do with the level of accessibility of a target noun at a lexical level: if a (subset) lexical relation can be established between a target noun and its antecedent, the acceptability rate of the target noun in preverbal position increases. We link this condition to a more general principle of D-linking, which, by hypothesis, is the same principle that can be used to explain various exceptional occurrences of weak indefinites in topic position. Thus, we suggest that D-linking principles facilitate the use of indefinite nominals in preverbal position, whether they are topics or not.

Olga Borik, Joan Borràs-Comes & Daria Seres

Acknowledgments

References

Baker, Mark C. 2015. *Case: Its principles and parameters*. Cambridge: Cambridge University Press.

Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and Meaning*, 119–162. Berlin: De Gruyter Mouton.

Beaver, David & Elisabeth Coppock. 2015. Novelty and familiarity for free. In Thomas Brochhagen, Floris Roelofsen & Nadine Theiler (eds.), *Proceedings of the 20th Amsterdam Colloquium*, 50–59. Amsterdam: ILLC/University of Amsterdam.

Borik, Olga. 2006. *Aspect and reference time*. Oxford: Oxford University Press.

Borik, Olga & Berit Gehrke. 2015. An introduction to the syntax and semantics of pseudo-incorporation. In Olga Borik & Berit Gehrke (eds.), *The syntax and semantics of pseudo-incorporation*, 1–43. Leiden/Boston: Brill.

Bryzgunova, Elena A. 1977. *Zvuki i intonacija russkoj reči*. Moskva: Russkij jazyk.

Carlson, Gregory. 1977. *Reference to kinds in English*. Amherst: University of Massachusetts. (Doctoral dissertation).

Cheng, Lisa & Rint Sybesma. 2014. Bare and not-so-bare nouns and the structure of NP. In C. T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 248–274. Wiley Blackwell.

Chierchia, Gennaro. 1984. *Topics in the syntax and semantics of infinitives and gerunds*. Arnhem: University of Massachusetts. (Doctoral dissertation).

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6(4). 339–405.

Christophersen, Paul. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Einar Munksgaard.

Chvany, Catherine V. 1973. On the role of presuppositions in Russian existential sentences. *PCLS* 9. 68–77.

Comrie, Bernard. 1981/1989. *Language universals and linguistic typology*. Chicago: The University of Chicago Press.

Comrie, Bernard. 1989. *The major languages of Western Europe*. London: Routledge.

Coppock, Elisabeth & David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* 38(5). 377–435.

Czardybon, Adrian, Oliver Hellwig & Wiebke Petersen. 2014. Statistical analysis of the interaction between word order and definiteness in Polish. In Adam Przepiórkowski & Maciej Ogrodniczuk (eds.), *Advances in natural language processing: 9th international conference on NLP (PolTAL 2014)*, vol. 8686 (LNCS/LNAI), 144–150. Springer.

Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4). 393–450.

Dimitrova-Vulchanova, Mila. 2012. Voice. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 937–959. Oxford: Oxford University Press.

Dryer, Matthew S. 2013. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at `http://wals.info/chapter/81`, Accessed on 2018-11-12).

Dyakonova, Marina. 2009. *A phase-based approach to Russian free word order*. Amsterdam: University of Amsterdam. (Doctoral dissertation).

Elbourne, Paul D. 2008. Demonstratives as individual concepts. *Linguistics and Philosophy* 31(4). 409–466.

Endriss, Cornelia. 2009. *Quantificational topics*. Dordrecht: Springer.

Erteschik-Shir, Nomi. 1997. *The dynamics of focus structure*. Cambridge: Cambridge University Press.

Farkas, Donka. 2002. Specificity distinctions. *Journal of Semantics* 19(3). 213–243.

Filip, Hana. 1999. *Aspect, situation types and noun phrase semantics*. New York/London: Garland Publishing Inc.

Frege, Gottlob. 1879. *Begriffsschrift: Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Verlag von Louis Nebert.

Fursenko, Diana I. 1970. Porjadok slov kak odno iz sredstv vyraženija neopredelënnosti / opredelënnosti imën suščestvitel'nyx. *Russkij jazyk za rubežom* 4. 68–72.

Galkina-Fedoruk, Evdokia M. 1963. *Vyraženije neopredelënnosti v russkom jazyke neopredelënnymi mesoimenijami i narečijami*. Moscow.

Geist, Ljudmila. 2008. Specificity as Referential Anchoring: Evidence from Russian. In Atle Grønn (ed.), *Proceedings of the 12th Sinn und Bedeutung conference (SuB12)*, 151–164. Oslo: ILOS.

Olga Borik, Joan Borràs-Comes & Daria Seres

Geist, Ljudmila. 2010. Bare singular NPs in argument positions: Restrictions on indefiniteness. *International Review of Pragmatics* 2(2). 191–227.

Hajičová, Eva. 1974. Meaning, presupposition, and allegation. *Philologica Pragensia* 17. 18–25.

Haspelmath, Martin. 1997. *Indefinite pronouns.* Oxford: Oxford University Press.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases.* Arnhem: University of Massachusetts. (Doctoral dissertation).

Heim, Irene. 2011. Definiteness and indefiniteness. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 1025–1058. Berlin: De Gruyter Mouton.

von Heusinger, Klaus. 1997. Salience and definiteness. *The Prague Bulletin of Mathematical Linguistics* 67. 5–23.

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

Isačenko, Alexander V. 1976. O grammaticheskom porjadke slov. In *Opera selecta*, 365–372. Munich: Fink.

Jasinskaja, Katja. 2014. Information structure in Slavic. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure.* Online Publication. DOI: 10.1093/oxfordhb/9780199642670.013.25.

Kamp, Hans. 1981. A theory of truth and semantic interpretation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Matematisch Centrum.

King, Tracy H. 1995. *Configuring topic and focus in Russian.* Stanford: CSLI Publications.

Krámský, Jiří. 1972. *The article and the concept of definiteness in language.* The Hague: Moutoun.

Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical matters*, 29–53. Stanford: CSLI Publications.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55. 243–276.

Krylov, Sergei. 1984. Determinacija imeni v russkom jazyke: teoretičeskije problemy. *Semiotika i informatika* 23. 124–154.

Leiss, Elisabeth. 2007. Covert (in)definiteness and aspect. In Werner Abraham, Elisabeth Stark & Elisabeth Leiss (eds.), *Introduction to nominal determination: Typology, context constraints and historical emergence*, 73–102. Amsterdam/Philadelphia: John Benjamins.

Leonetti, Manuel. 2010. *Specificity and topicality: Some thoughts on their complex interaction.* Talk given at the workshop 'Specificity from empirical and theoretical points of view', Stuttgart University.

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

Mathesius, Vilém. 1964. On some problems of the systematic analysis of grammar. In Josef Vachek (ed.), *A Prague School reader in linguistics*, 306–319. Bloomington: Indiana University Press.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60. 847–894.

Nesset, Tore. 1999. The realization of (in)definiteness in Russian. *Poljarnyj Vestnik* 2. 84–109.

Partee, Barbara. 1987. Noun-phrase interpretation and type-shifting principles. In Jeroen Groenendijk & Martin Stokhof (eds.), *Studies in discourse representation theory and the theory of generalized quantifiers, GRASS 8*, 115–143. Dordrecht: Foris.

Pereltsvaig, Asya. 2000. Monotonicity-based vs. veridicality-based approaches to negative polarity: Evidence from Russian. *Formal Approaches to Slavic Linguistics* 8. 328–346.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland & Alice G. B. ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, MA: MIT Press.

Portner, Paul & Katsuhiko Yabushita. 2001. Specific indefinites and the information structure theory of topics. *Journal of Semantics* 18(3). 271–297.

Pospelov, Nikolaj S. 1970. O sintaksičeskom vyraženii kategorii opredelënnosti – neopredelënnosti v sovremennom russkom jazyke. In *Issledovanija po sovremennomu russkomu jazyku*, 182–189. Moskva.

Reinhart, Tanya. 1981. Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27(1). 53–94.

Rochemont, Michael. 2016. Givenness. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 41–63. Oxford: Oxford University Press.

Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493.

Schoorlemmer, Maaike. 1995. *Participial passive and aspect in Russian.* Utrecht: Utrecht University. (Doctoral dissertation).

Sgall, Petr. 1972. Topic, focus, and the ordering of elements of semantic representations. *Philologica Pragensia* 15. 1–14.

Olga Borik, Joan Borràs-Comes & Daria Seres

Šimík, Radek. 2014. Definiteness and articleless languages. Unpublished ms. (Available online at https://amor.cms.hu-berlin.de/ simikrad/pdf/simik-dacice-definiteness.pdf, Accessed on 2018-19-11).

Šimík, Radek & Markéta Burianová. 2018. Definiteness of bare NPs as a function of clausal position: A corpus study of Czech [preprint]. In *Proceedings of FASL 26*. Michigan Slavic Publications.

Šimík, Radek & Marta Wierzba. 2015. The role of givenness, presupposition, and prosody in Czech word order: An experimental study. *Semantics and Pragmatics* 8(3). 1–103.

Strawson, Peter F. 1950. On referring. *Mind* 59. 320–344.

Szwedek, Alexander. 1974. Some aspects of definiteness and indefiniteness of nouns in Polish. *Papers and studies in contrastive linguistics* 2. 203–211.

Topolinjska, Zuzanna. 2009. Definiteness (synchrony). In Sebastian Kempgen, Peter Kosta, Tilman Berger & Karl Gutschmidt (eds.), *The Slavic languages*, 176–187. Berlin/New York: De Gruyter.

Verkuyl, Henk J. 1998. Tense, aspect and aspectual composition. *UiL Working Papers (February 1998)*.

Wierzbicka, Anna. 1967. On the semantics of verbal aspect in Polish. In *To honor Roman Jakobson: Essays on the occasion of his seventieth birthday*, 2231–2249. The Hague/Paris: Mouton.

Wolter, Lynsey. 2004. Demonstratives, definiteness and determined reference. In Keir Moulton & Matthew Wolf (eds.), *NELS 34: Proceedings of the 34th annual meeting of the North East Linguistic Society*, 603–617. Amherst, MA: GLSA Publication.

Yanovich, Igor. 2005. Choice-functional series of indefinite pronouns and Hamblin semantics. In Efthymia Georgala & Jonathan Howell (eds.), *Proceedings of the 15th Semantics and Linguistic Theory Conference*, 309–326. Ithaca: Cornell University.

Yokoyama, Olga. 1986. *Discourse and word order*. Amsterdam/Philadelphia: John Benjamins.

Chapter 4

Referential anchoring without a definite article: The case of Mopan (Mayan)

Eve Danziger
University of Virginia

Ellen Contini-Morava
University of Virginia

Mopan Maya is a language in which pragmatic factors play a significant role in referential anchoring. Its article occurs in both definite and indefinite contexts, and so do bare nominals. We discuss several forms that assist with referential anchoring, using Dryer's (2014) reference hierarchy as an organizing framework, but none of these forms is obligatory for any of the functions in the hierarchy. Rather than explicitly encoding, e. g., definiteness or specificity, their employment is sensitive to factors such as discourse salience.

1 Introduction

It is now well documented (e. g., Sasse 1988; Matthewson 1998; Gillon 2009; 2013; Davis et al. 2014: e201-e207; Lyon 2015) that languages exist in which determiners do not signal semantic gradations of relative 'definiteness' (degrees of identifiability and uniqueness, see Hawkins 1978; Löbner 1985; 2011; Lyons 1999; Dryer 2014), as they do in most European languages (for example, by the contrast between English *the* and *a/an*). The question therefore arises whether degrees of definiteness are explicitly signaled in such languages, or if not, how related messages can be conveyed. In the following, we discuss the case of Mopan Mayan (Yukatekan), a language in which the form that fills distributional criteria to be an article does not encode the semantic concept of definiteness or the related concepts of specificity and uniqueness (Contini-Morava & Danziger forthcoming). We address the

various means which are used in Mopan to indicate relative identifiability and uniqueness, using a scale developed by Dryer (2014) specifically to handle the notions 'definite/indefinite' in typological comparison.[1] We describe a number of forms which can be used to indicate relatively high or low degrees of identifiability, and we also document the fact that ART can occur at every position on Dryer's hierarchy, thus confirming that ART does not usefully convey information about identifiability or uniqueness.

We also note however that in very many Mopan discourse cases, explicit means of indicating the status of a referent vis-à-vis identifiability and uniqueness are not in fact employed. We show that in all positions on Dryer's hierarchy, referents can be expressed by unmarked, or 'bare', nominals,[2] and therefore no explicit information is provided about degrees of identifiability or uniqueness. We conclude that the identification of referents in terms of degrees of previous mention, uniqueness, specificity, or familiarity to speech participants is not always explicitly formulated in Mopan. When this is the case, calculation of these properties of referents must be accomplished, if it is accomplished at all, by pragmatic means.

1.1 Resources for referential anchoring in Mopan

In Mopan, information about the referential status of argument expressions can be provided in a variety of ways. These include:

(a) use of the article (ART), which explicitly signals that the associated constituent is to be construed as an entity, and hence an argument;

(b) use of a demonstrative expression involving one of four Stative deictic predicates that specify proximity, visibility, and states of prior knowledge to various speech-act participants and discourse referents (Danziger 1994), with or without an accompanying NP;

(c) use of the emphatic 3rd person pronoun *le'ek*;

---

[1]Other scales and metalanguages, such as Gundel et al.'s (1993) givenness hierarchy or Löbner's (2011) uniqueness scale, would have been reasonable alternatives. For present purposes, Dryer's hierarchy has the advantage in that it pursues degrees of 'indefiniteness' as well as of 'definiteness', and does not deal with contrasts other than those of 'definiteness'. (For example, 'relationality' is not a dimension on Dryer's scale.)

[2]The distinction between 'noun' and 'verb' as separate lexical classes is problematic in Mopan (Danziger 2008; see further below), but for ease of reference we will use the terms 'noun' and 'nominal' to mean 'word understood as serving in a given utterance as an argument in the predication, as possessor in a possessive phrase, or as object of a preposition', and 'verb' to mean 'word understood as serving as predicator in a given utterance'.

(d) use of a numeral classifier phrase;

(e) a bare nominal: absence of any explicit specification of referential status.

The chapter is organized as follows. In section 2, we provide distributional and semantic characterization of the above forms. In section 3, we introduce Dryer's (2014) reference hierarchy, which we use as a framework for fuller description of the 'definiteness/indefiniteness' functions of the forms listed above. We show how some of the listed means of expression are restricted to certain portions of the scale, meaning that they can be characterized as conveying degrees of definiteness or indefiniteness. But we also show that both ART and 'bare nominal' can occur in any position on the hierarchy. This means on the one hand that ART does not usefully convey degrees of 'definiteness', and on the other that none of the definiteness-conveying means which we also document is actually required, even when its preferred segment of the scale is at issue in a given utterance. That is, it is often the case that degrees of, e. g., anaphora, specificity, etc. are pragmatically inferred rather than semantically conveyed by reliance on dedicated grammatical forms. Section 4 provides a summary and conclusion. A table summarizing the data appears in this final section.

Our data are drawn primarily from Mopan narratives, including 75 narratives from eight speakers collected by Pierre Ventur in Guatemala in the 1970s (Ventur 1976), 14 texts of varied kinds from ten speakers collected by Matthew and Rosemary Ulrich (Ulrich & Ulrich 1982), and narratives collected in Belize more recently by Eve Danziger (p.c.) and by Lieve Verbeeck (Verbeeck 1999). We also draw on conversational data elicited by Eve Danziger from Mopan speakers in Belize (Danziger 1994).

2 Descriptive preliminaries

Mopan is a Mayan language spoken by several thousand people living in communities that span the Belize-Guatemala border in Eastern Central America. It is a predominantly head-marking, predicate-initial language. Pluralization is optional, numeral classifiers are required for enumeration, and there is no copular verb.

2.1 The article (ART) *a*

The Mopan lexicon is characterized by the neutrality of many lexical items in relation to the traditional distinction between noun and verb (Danziger 2008). Many lexical items which would translate into English as nouns may play the role of

a clause predicate without derivation. Such items fall into the category of 'statives' in Mopan (see Danziger 1996; for similar observations in other Yukatekan languages see Bricker 1981; Lucy 1994; Lois & Vapnarsky 2006). This can be seen in example (1a), where the lexeme *winik*, inflected with the pronominal suffix from the series known to Mayanists as Set B, is interpreted as a stative predicate ('be a man'). In other contexts, such as (1b), the same lexeme is construed as an argument ('the man' or 'a man'). Note the presence of ART *a* before *winik* in example (1b).

(1) Noun-verb neutrality in Mopan.

  a. Stative lexeme with 2nd person Set B inflection.[3]
     [Author's data, *Ix Che'il etel Bäk'* 'Wild Woman', J. S.]

     inchech=e tan-Ø   inw   il-ik-**ech**.
     2.EMPH=EV be_continuing-3B 1A(prevocalic) see-TR.IPFV-**2B**

     winik-**ech**.
     man-**2B**

     'As for you, I am looking at **you**. **You're (a) man**.'

  b. Same lexeme with ART.
     [Source as in (1a)]

     "..." kut'an **a**   **winik** t-uy   ätan=a.
        3.QUOT ART man   PREP-3A(prevocalic) wife=EV

     ' "..." said **the [ART] man** to his wife.'

ART designates an entity that instantiates the content of the accompanying constituent (see Contini-Morava & Danziger forthcoming for details). As such, it helps to distinguish arguments unambiguously from predicates.

We show below that ART does not usefully convey semantic contrasts on the definiteness dimension. If this is the case, are we justified in calling it an 'article'? Although some have argued that semantic criteria such as definiteness or specificity are necessary for defining the category 'article' (e. g., Himmelmann 2008: 833-4), others foreground distributional criteria. For example, Dryer (2007: 158) states that the term 'article' can be applied to "a set of words which occur with high frequency in noun phrases and which vary for certain grammatical features

---

[3]Orthography is as preferred by the Academía de las Lenguas Mayas de Guatemala (ALMG, England & Elliott 1990). Interlinear glosses follow the Leipzig glossing rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php), with some additions; see Abbreviations at the end of the chapter.

of the noun phrase".[4] Mopan ART occurs in a fixed position preceding expressions that are to be construed as nominals. It is also in complementary distribution with forms that function as possessive pronouns (the pronoun series known to Mayanists as Set A), again suggesting determiner status.[5] ART is glossed as 'the' in several previous works on Mopan (Shaw 1971; Ulrich & Ulrich 1982; Ulrich et al. 1986), even though it can be used in contexts that cannot be construed as definite (see section 3 below); Hofling (2006) glosses it as DET[erminer]. We use the term 'article' to distinguish *a* from the Set A possessives with which it is in complementary distribution and which might also be considered to be DET[erminer].

Aside from occurring before single lexical items as in example (1b), ART also occurs before 'property concepts' and other expressions, if they are to function as arguments, as in (2).

(2) ART preceding lexemes usually construed as adjectival modifiers. [Ventur (1976) 3:16, *Aj Känän Kax* 'The Chicken Keeper', E. S.][6]

jok'-ij　　　　**a　nooch=o**. Tal-ij　　　　　　a　**nene'=e**.
exit-3B.INTR.PFV ART big=EV　come-3B.INTR.PFV ART small=EV
'**The [ART] big (one)** left off (lit. went out). **The/a [ART] little (one)** came.'

In some cases, ART's ability to allow forms that do not normally denote entities to function as clause arguments yields an English translation as a relative clause. In example (3), the article precedes something that would otherwise be interpreted as a predicate 'he went under the bed').[7]

---

[4] In his WALS study Dryer defines 'articles' more narrowly as "words or morphemes that occur in noun phrases…[that] must code something in the general semantic domain of definiteness or indefiniteness" (2014: e234), but this was for the purpose of surveys specifically of definite and indefinite articles.

[5] A possessive construction involves two referents, each of which may require its own referential anchoring. As such, they do not fit easily into Dryer's (2014) reference hierarchy, used below as an organizing framework for our discussion. Dryer suggests that a possessor is inherently an indication of an NP's definiteness (fn 4, p. e234), and he does not include possessive constructions in his discussion. Others however (e. g., Alexiadou 2005) argue that possessives are not always definite. Given the complexity of integrating possessive constructions with Dryer's hierarchy, we will not discuss them further here.

[6] Ventur's collection of narratives, transcribed and translated into Spanish by Ventur and his Mopan consultants, was donated by Ventur to the Smithsonian. We provide our own interlinear glosses and translations into English. Ventur's manuscript includes the names of the original narrators, but since we have no way to obtain permission to publish their names, we use only initials to refer to them. For examples from published sources we include the full names of the speakers.

[7] A reviewer asks why we do not just use the gloss 'nominalizer' for ART. One reason is its complementary distribution with the possessive pronouns, mentioned above as a criterion for

(3) ART preceding predicative expression.
     [Ventur (1976) 1:03, *Aj Okol ich Witz*, 'He who Enters the Mountain', R. K'.]

"..." kut'an b'in **a**   **b'in-ij**   **yalan kamaj=a**.
    3.QUOT HSY ART go.PFV-3B.INTR.PFV under bed=EV

' "..." said **the [ART] (one who) had gone under the bed**.'

2.1.1 ART with relativized deictic predicates

One frequent example of the relativizing function of ART that will be relevant
to our discussion of referential anchoring is its use with a set of four dedicated
stative deictic predicates that provide information about referents with respect
to their proximity, visibility, and states of prior knowledge to various speech-
act participants. These are *la'~d'a'* 'deictic stative 1st person', *kan(a')* 'deictic
stative 2nd person', *lo'~d'o'* 'deictic stative 3rd person known through visual
means', and *b'e'* 'deictic stative 3rd person known through other than visual
means' (Danziger 1994). When a predicate of this series is relativized using ART
a, the result is a form most simply rendered in English as a deictic demonstrative
('this one/that one'). A more literal translation recognizes the predicate content,
and might read 'one who/which is near me', 'one who/which is near you', etc.
(Danziger 1994: 891-894, see also Jelinek 1995: 489-490 for similar analysis of De-
terminer Phrases in Straits Salish). We therefore refer to these demonstrative ex-
pressions as 'relativized deictics'. We do not include the deictic predicates *do'~lo'*
'deictic stative 3rd person visible', *da'~la'* 'deictic stative 1st person', and *kan(a')*
'deictic stative 2nd person' in the discussion which follows, because these forms
are used primarily in face-to-face conversation, and the categories of Dryer's
hierarchy are better suited for application to narrative contexts.

A relativized deictic can occur alone or together with lexical specification
of the referent. (4) is an example of the latter.

---

determiner status. Another is that lexemes can function as 'nominals' (clause arguments) in
Mopan with or without ART (see example (8) below).

(4) Lexeme with relativized deictic expression.
[Ventur (1976) 5:07, *Aj Ma' Na'oo'* 'The Orphans', J. I.]

pero **a   winik a   b'e=e,**   u ka'   käx-t-aj-Ø
DISC ART man   ART D.3.NV=EV 3A again seek-T-TR.PFV-3B

u laak' uy   ätan.
3A other 3A(prevocalic) wife

'So **that [ART] man**, **known by other than visual means**, he looked again
for another wife.'

As we will show below, a relativized deictic may be employed to indicate
identifiability of a discourse referent.

2.1.2 Emphatic pronoun

Mopan is a polysynthetic language in which verb arguments are frequently en-
coded only in obligatory person affixes of the verb. (See for instance the 2nd
person Set B affix in example (1a), 'you are a man'.) This includes arguments
which denote referents previously mentioned in the discourse.

(5) 3rd person undergoer affix for anaphoric reference.[8]
[Ventur (1976) 3:15, *Siete Kolor* 'Seven Colors', E. S.]

sas-aj-ij   samal-il=i,
lighten-INCH-3B.INTR.PFV next.day-POSS=EV

ka' b'in-oo'   tukadye'   u käx-t-aj-Ø-oo'   b'in.
again go.PFV-3.PL another.time 3A seek-T-TR.PFV-3B-3.PL HSY

'(When) it dawned the next day, they went and looked for **it** again.'

In this example 'it' (expressed by the zero Set B suffix) refers to previously
mentioned coffee and cacao for the king's horse to eat, after the protagonists
have been unsuccessful in finding this food the day before. The unusual food has
already been named and discussed at length.

If emphasis on a particular argument is desired, it is possible to add a
person-indicating independent pronoun. The third person in this series has the
form *le'ek* and is relevant to our discussion of referential anchoring. *Le'ek* oc-
curs twice in the example below, which comes from a story in which a young
woman's father has shot a small hummingbird which he found in her bedroom,

---

[8]The 3rd person Set B undergoer affix is a zero morpheme.

and now comes to understand that this hummingbird was actually a magical disguise for his daughter's lover, the Holy Sun. The first use of *le'ek* ('that hummingbird I shot') occurs in combination with a nominal (*tz'unu'un*, 'hummingbird') and helps to specify which hummingbird we are talking about. The second use ('that was the Holy Sun') occurs alone as one side of an equational predication.

(6)   *Le'ek*, emphatic pronoun.
     [Ventur (1976) 1:05, *U kwentojil Santo K'in y Santo Uj* 'The Story of the Holy Sun and Holy Moon', R. K'.]

     le'ek   **a**   **tz'unu'un**   in tz'on-aj-Ø=a,
     3.EMPH ART hummingbird 1A shoot-TR.PFV-3B=EV

     le'ek   a   santo k'in=i.
     3.EMPH ART holy   sun=EV

     '**That hummingbird** I shot, **that** was the Holy Sun!' [Lit. That which is a hummingbird I shot, is that which is the Holy Sun!]

2.1.3 Numeral + classifier construction

In Mopan, enumeration of nominals requires use of a numeral classifier. A numeral classifier phrase consists of numeral + classifier (+ optional ART) + nominal. It is overwhelmingly the numeral *jun* 'one' that is found in this function, although other numerals can also introduce referents where appropriate. This construction is often used to introduce new discourse referents.[9] An example is (7), the first sentence in a story; see also section 3.3.4 below.

(7)   Numeral classifier construction.
     [Ventur (1976) 1:08, *Aj Jook'* 'The Fisherman', R. K'.]

     jun tuul   b'in **a**   **winik=i**,
     one CLF.ANIM HSY ART man=EV

     top   ki'-Ø   b'in t-u   wich a   jook'=o.
     very be.good-3B HSY PREP-3A eye   ART fishing=EV

     '**A man**, fishing was very good in his eye(s) (he liked fishing very much).'

2.1.4 Bare nominal

Despite the abovementioned noun-verb lexical fluidity that is characteristic of Mopan, it is possible for a bare lexical item to be construed as an argument if its

---

[9]Use of the numeral 'one' for discourse-new referents is common cross-linguistically and is often the source for indefinite articles (see, e. g., Lyons 1999).

lexical meaning readily supports this. An example is (8).

(8)   Bare lexical item interpreted as argument.
      [Author's data, *Ix Che'il etel Bäk'* 'Wild Woman', J. S.]

      o, inen=e   waye' watak-en
      oh 1.EMPH=EV D.LOC.1 be.imminent-1B

      waye' yan-Ø   in kaal,   kut'an **winik=i**.
      D.LOC.1 exist-3B 1A hometown 3.QUOT man=EV

      ' "Oh, myself, I come from here. Here [this] is my home village,"
      said **(the) man**.'

    Here the word *winik* '(be a) man' follows a direct quotation, along with
the quotative *kut'an*, so it is readily interpreted as the one doing the saying, i. e.,
as an argument. We will show that bare nominals may be ascribed a wide range
of definiteness interpretations in Mopan.

3 Dryer's (2014) reference hierarchy

As an organizing framework for discussing anchoring, we will use the reference
hierarchy described by Dryer (2014: e235), the basis for his chapter on definite ar-
ticles in the World Atlas of Language Structures (https://wals.info/chapter/37).
Dryer proposes that a hierarchical organization facilitates cross-linguistic com-
parison, and asserts that any article which accomplishes the leftmost functions
in the hierarchy, to the exclusion of at least some functions on the right, should
be declared a definite one (Dryer 2014: e241).[10] Dryer's hierarchy was intended

---

[10]Dryer (2014: e237-238) treats preferential occurrence of an article on a contiguous span of his
   reference hierarchy as the basis for classifying the article as 'definite' or 'indefinite', depending
   on whether its span is located toward the left or right of the hierarchy. He classifies the Basque
   article as 'definite' even though it occurs in all positions of the hierarchy (Dryer 2014: e239),
   because it cannot occur in a subset of indefinite contexts (semantically nonspecific indefinites
   within the scope of negation). This may be an acceptable heuristic for typological purposes (or
   it may not, see Contini-Morava & Danziger forthcoming), but it does not solve the potential
   semantic ambiguity of actual occurrences of Mopan ART as regards identifiability or unique-
   ness, when this form occurs in actual discourse. In fact ART can occur within the scope of
   negation, as in the following example, uttered by an unsuccessful hunter:

      ma' yan-Ø a   b'äk=a.
      NEG exist-3B ART game=EV
      'There isn't any game.'
      [Author's data, *Ix Che'il etel Bäk'* 'Wild Woman and Meat', J. S.]

for typological comparison specifically of articles, but we include a broader set of anchoring devices in order to provide a fuller picture of referential anchoring in Mopan. The typological aspects of Dryer's proposal are of less interest to us here than the usefulness of his framework for descriptive organization in a single language. His hierarchy is as follows:

Dryer's reference hierarchy (Dryer 2014: e235)[11]
anaphoric definites > nonanaphoric definites > pragmatically specific indefinites > pragmatically nonspecific (but semantically specific) indefinites > semantically nonspecific indefinites

A brief explanation of terms that may not be familiar to the reader (see Dryer 2014: e236-e237): An anaphoric definite NP refers back in the discourse, i. e., is "licensed by a linguistic antecedent" (Dryer 2014: e236), whereas a non-anaphoric definite relies instead on shared knowledge between speaker and addressee; an example of the latter would be *the sun* (in a context where there are not multiple suns). These notions of definiteness have much in common with prior understandings (e. g., Hawkins 1978; Lyons 1999), that definiteness is a matter of encoding 'identifiability' and/or 'inclusivity' (more on these ideas below). It is useful for our purposes, however, that Dryer's hierarchy also extends to characterization of the semantics of indefinites.

For Dryer, semantically specific indefinites are those where there is an entailment of existence (e. g., *I went to a movie last night*). Within this type, Dryer distinguishes between pragmatically specific indefinites which indicate a discourse participant that "normally ... is referred to again in the subsequent discourse" (Dryer 2014: e236), and pragmatically nonspecific indefinites (an NP whose referent is not mentioned again, even though there is an entailment of existence).

Finally, a semantically nonspecific indefinite NP (which necessarily is also pragmatically nonspecific) does not entail existence of the referent, e. g., *John is looking for a new house*.[12]

In the following, we document the distribution of the Mopan forms described above across each of the positions of Dryer's hierarchy. One of our principal findings is the fact that the 'bare nominal' option is allowable across all po-

---

[11]Dryer (2014: e235) states that his hierarchy is based on Givón's (1978) 'wheel of reference', but Dryer uses some different terminology and omits generics and predicate nominals from his hierarchy.

[12]Dryer (2014: e237) acknowledges that a semantically nonspecific referent can be mentioned again (i. e., could fit his definition of 'pragmatically specific'), as in *John is looking for a new house. It must be in the city...* He also states, however, that "articles that code pragmatic specificity appear never to occur with semantically nonspecific noun phrases" (ibid.). He does not include the category of semantically nonspecific but pragmatically specific in his hierarchy.

sitions in the hierarchy. This means that, even if other forms can be said (based on their distribution across the hierarchy) to encode definite or indefinite semantics, these forms are never obligatory in the relevant semantic contexts. In many cases, therefore, it seems that distinctions of referential anchoring in Mopan are made pragmatically, based on context.

We also make special note of the fact that, while it is never obligatory, Mopan ART is allowable in all positions on the hierarchy. ART, therefore, cannot be said to encode any sort of distinction between the semantic positions in the hierarchy (that is, it does not encode any semantics of definiteness).

We now consider each position on Dryer's (2014) hierarchy in turn, describing the central Mopan possibilities in each case.

3.1 Anaphoric definites

3.1.1 ART

In Mopan, anaphoric definites are frequently preceded by ART. In (9), the referent has been mentioned in the immediately preceding context and is known to both the storyteller and the addressee.

(9) ART in contexts consistent with anaphoric definiteness.
     [Author's data, *Ix Che'il etel Bäk'* 'Wild Woman', J. S.]

     "..." kut'an **a**   **winik** t-uy   ätan=a.
        3.QUOT ART man   PREP-3A(prevocalic) wife=EV
     ' "..." said **the [ART] man** to his wife.'

We will show, however, that ART does not explicitly encode anaphoric definiteness, since it can also be found in nonanaphoric and non-definite contexts (see sections 3.2-3.5 below).

3.1.2 ART + deictic predicate

More explicit indication of anaphoric definiteness may also be accomplished through the use of ART to create a relative clause from the deictic predicate *b'e'* 'near neither speaker nor hearer and known through non-visual means'. The non-visual means in question are commonly understood to include prior mention in discourse (Danziger 1994). This construction therefore yields an expression that is equivalent to an anaphoric deictic demonstrative. This was shown in example (4), repeated as (10) for convenience.

Eve Danziger & Ellen Contini-Morava

(10)   Deictic expression for anaphoric definite.
       [Ventur (1976) 5:07, *Aj Ma' Na'oo'* 'The Orphans', J. I.]

       pero **a   winik a   b'e=e,**   u ka'   käx-t-aj-Ø
       DISC ART man   ART D.3.NV=EV 3A again seek-T-TR.PFV-3B

       u laak' uy   ätan.
       3A other 3A(prevocalic) wife

       'So **that [ART] man**, **known through non-visual means**, he looked again
       for another wife.'

The predicate *b'e'* can itself occur alone with ART, yielding a referential ex-
pression translatable as 'one which is near neither speaker nor hearer and which
is known through non-visual means', as in (11).

(11)   Anaphoric definite with relativized deictic predicate *a b'e'* 'deictic stative
       3rd person non-visible' used alone.
       [Ventur (1976) 3:11, *Uj y k'in* 'Moon and Sun', E. S.]

       top kich'pan-Ø   ti   in wich,
       very be.beautiful-3B PREP 1A eye

       kut'an b'in **a   b'e'=e.**
       3.QUOT HSY ART D.3.NV=EV

       ' "I like it very much," said **that one known through non-visual means**.'

Example (11) comes in the middle of a story in which a young woman (the
Moon) has been speaking to her father. In the preceding context her quotations
are interspersed with the expression *k'u t'an b'in* 'apparently [that is] what [s/he]
said', which is very common for quotations in Mopan narrative, and completely
lacks overt identification of the speaker. This example comes at the end of her
conversational turn, just before her father's reply. Although it has been clear all
along who the speaker is, here the narrator makes the anaphoric reference more
explicit by means of the deictic, perhaps to mark the transition to a new speaker.
In any case, no lexical specification is needed, and the deictic is used alone.

3.1.2.1 Optionality of relativized deictic for explicit marking of anaphoric defi-
niteness

Recall that in the context immediately preceding example (11) above there are
several non-explicit allusions to the woman being quoted, in contrast with the
deictic expression that appears in the cited example. This example thus illustrates

another characteristic of referential anchoring of anaphoric definites in Mopan: even though this information can be conveyed by a deictic expression, a deictic is not obligatory with anaphoric definites. This is shown in (12), in which there are two anaphoric NPs, but only the second one is marked by a deictic.

(12)   Anaphoric definites with and without relativized deictic predicate.
       [Author's data, *Ix Che'il etel Bäk'* 'Wild Woman and Meat', J. S.]

       ma' patal-Ø   u   ch'uy-t-e'   a   b'äk'=ä
       NEG be.able-3B 3A hang-T-TR.IRR.3B ART meat=EV

       a   winik a   b'e'.
       ART man   ART D.3.NV

       'That [ART] man, **known through non-visual means**, couldn't hoist up
       the [ART] meat.'

In example (12), the protagonist has been mentioned several times, and has encountered a supernatural forest woman, who has brought him a large quantity of game. The game is so heavy that the man can't lift it to take it home. Here there are two anaphoric NPs: *a b'äk'* 'the meat' and *a winik a b'e'* 'that man'. The first is marked only by ART and the second by both ART and a relativized deictic.

One could ask why the deictic is used in (12) at all, since this is the only man mentioned in the story so far. Why not just use ART + nominal, as is done with the reference to the meat (also previously mentioned in the story)? In this case the deictic appears to add emphasis: in contrast with the woman, who had no trouble carrying the meat, and in contrast to other possible men who might also be able to carry it, that particular man was unable to lift it.[13]

When referring to anaphoric definites, then, a relativized deictic can be used, but is not obligatory. It is also allowable, and far from unusual, for ART alone to occur in such contexts. There may be a tendency for relativized deictics to be associated with contrast or extra emphasis, but further research would be needed to confirm this.

3.1.3 The emphatic pronoun *le'ek*

Le'ek 'be it/be the one' is appropriately used for anaphoric mention, as in (13).

---

[13]This interpretation is also consistent with the use of the deictic in example (11), where the deictic marks a transition between speakers.

Eve Danziger & Ellen Contini-Morava

(13) *Le'ek* for anaphoric mention.
[Ventur (1976) 3:15, *Siete Kolor* 'Seven Colors', E. S.]

käkäj i   kafe,
cacao and coffee

le'ek   a   walak-Ø   u jan-t-ik-Ø   in kabayoj=o.
3.EMPH ART be.habitual-3B 3A eat-T-TR.IPFV-3B 1A horse=EV

'Cacao and coffee, **it is that** which my horse eats.'

Relativized deictics, including *b'e'* 'associated with neither speaker nor hearer and known through non-visual means' can occur with *le'ek*, as shown in (14).

(14) *Le'ek* with relativized deictic *a b'e'*.
[Ventur (1976) 3:15, *Siete Kolor* 'Seven Colors', E. S.]

le'ek   a   b'e'   u p'o'-aj-Ø=a.
3.EMPH ART D.3.NV 3A do.laundry-TR.PFV-3B=EV

'**It is he**, **known through non-visual means**, who washed the clothes.'

In this story, the hero has been secretly out winning the competition to marry the princess, but now returns home to the humble identity of a hard-working younger brother, assigned to menial domestic tasks.

Finally, *le'ek* can also co-occur in anaphoric use with a nominal phrase with ART plus a relativized deictic, as shown in (19).[14]

---

[14]This construction, applied to each of the deictic predicates in turn, is cognate with the current Yukatek demonstrative series (Hanks 1990).

4 *Referential anchoring without a definite article: The case of Mopan (Mayan)*

(15) *Le'ek* + ART + nominal + relativized deictic.
[Ventur (1976) 5:06, *Kompadre etel a Komadre* 'The Compadre and the Co-madre', J. I.]

tz'a'-b'-ij   u meyaj
give-PASS-3B.INTR.PFV 3A work

ichil jum p'eel   jardin. ...
inside one CLF.INAN garden

bueno. **le'ek** **a** **meyaj a** **b'e'** u b'et-aj-Ø=a
well   3.EMPH ART work ART D.3.NV 3A do-TR.PFV-3B=EV

'He was given work in a garden. ... well, that work is what he did.'
[Lit. Well, **it is that which is work which is known through non-visual means** (that) he did]

    In addition to serving as a stative predicate, the emphatic pronoun, then, may also appear with nominals, and it is an important resource in Mopan for indicating reference to a previously mentioned referent. As we have shown, however, *le'ek* is not obligatory for anaphoric reference.

3.1.4 Bare nominal

In Mopan, it is possible for a bare nominal to be used for anaphoric definite reference, as shown in (8), repeated for convenience as (16).

(16) Bare nominal for anaphoric definite referent.
[Author's data, *Ix Che'il etel Bäk'* 'Wild Woman', J. S.]

o, inen=e   waye' watak-en
oh 1.EMPH=EV D.LOC.1 be.imminent-1B

waye' yan-Ø in kaal,   kut'an **winik=i**.
D.LOC.1 exist-3B 1A hometown 3.QUOT man=EV

' "Oh, myself, I come from here. Here [this] is my home village,"
said **(the) man**.'

    Here the man is the main protagonist in the story, and has been mentioned several times before. As mentioned earlier, use of a bare referring expression occurs only when its lexical semantics support argument construal (see Contini-Morava & Danziger forthcoming for details).

Eve Danziger & Ellen Contini-Morava

3.2 Nonanaphoric definites

3.2.1 ART for nonanaphoric definites

Dryer (2014: e236) defines nonanaphoric definites as definite noun phrases whose use "is based only on shared knowledge of the speaker and hearer", unlike anaphoric definites whose use is "licensed by linguistic antecedents" (ibid.). With regard to prior mention, Dryer further states that "in English, one would not normally refer to the sun with the noun phrase *the aforementioned sun*, even if there were a previous reference to it" (ibid.), presumably because *the sun* has a unique referent, so does not require re-identification. Mopan ART occurs readily in such contexts, as shown in example (17).

(17) ART for unique individuals.
      [Ulrich & Ulrich (1982), 'Mopan Maya Concept of Earth and Heaven', José María Cowoj, interviewed by Matthew Ulrich, line 41][15]

      a   uj=u   tan-Ø   ilik u b'eel
      ART moon=EV be.continuing-3B INT 3A go.IPFV

      jab'ix ti   tan-Ø   u b'eel   **a** **k'in**.
      like   PREP be.continuing-3B 3A go.IPFV ART sun
      '**The [ART] moon** goes along just like **the [ART] sun** goes.'

    In (17), the moon and the sun could be construed as definite because each has a unique referent.[16]

3.2.2 Relativized deictic for nonanaphoric reference

Although the relativized deictic *a b'e'* is most often used for anaphoric reference, it can also occur with unique referents, as in example (18).

---

[15]For all examples from this source, we regularize the orthography to that recommended by the ALMG (see supra note 3) and provide our own glossing.

[16]Löbner (2011: 282) treats e. g., *moon* as an 'individual noun', marked by the feature [+Unique], i. e., as 'semantically definite' and inherently unique. By contrast, a noun like *man* is a 'sortal noun', i. e., [–Unique], but it can be coerced into an individual reading by contextual information that identifies a particular individual, which can make it 'pragmatically definite' in a given context (pp. 307-308). We will see below (examples (30) and (32)) that Mopan ART does not coerce an individual reading for the associated nominal.

(18) Relativized deictic for nonanaphoric definite.
[Ventur (1976) 3:07, *U Kweentojil aj Peedro* 'The Story of Pedro', E. S.]

ok-ij   b'in ichil **a   ka'an a   b'e'=e**.
enter-3B.INTR.IPFV HSY inside ART sky   ART D.3.NV=EV
'He went inside **that sky**.'

Example (18) is from a story in which a man wants to enter the sky in order to see God, and he is finally allowed in after a series of negotiations with Saint Peter. Like the sun and moon in (17), the sky is unique, so even though the sky has been mentioned before in this story, according to Dryer's definition, example (18) would not constitute anaphoric reference. The relativized deictic is not being used in order to remind the hearer that we are talking about the same sky that has been mentioned before. In this example it seems merely to add emphasis (cf. example (12), discussed earlier).

3.2.3 Emphatic pronoun *le'ek* 'be a 3rd person' for nonanaphoric reference

The independent pronoun *le'ek* 'be a third person' may also be used for inherently unique referents. In example (19), from a legend in which Jesus is hunted down by evil pursuers, first mention of this very familiar and unique protagonist is made using *le'ek*.[17] This usage helps to specify that we are talking about a unique referent rather than just one man among others who bears this name.

(19) *Le'ek* for nonanaphoric definite.
[Ventur (1976) 5:09, *U Alkab'eeb' Jesus* 'The Chasing of Jesus', J. I.]

bueno, **le'ek   a   jesus=u**,
well   3.EMPH ART Jesus=EV

ti   kaj-ij   alka'-b'-äl
PREP begin-3B.INTR.IPFV run-PASS-INTR.IPFV
'Well, **he who is Jesus**, when he was beginning to be chased, …'

3.2.4 Bare nominal for nonanaphoric reference

It is also possible for a bare nominal to be used for nonanaphoric definite reference, as in (20).

---

[17]Jesus is also introduced with ART, rather than with the masculine gender marker, which would normally be expected with the name of an ordinary human man (Contini-Morava & Danziger 2018).

Eve Danziger & Ellen Contini-Morava

(20) Bare nominal for nonanaphoric definite reference.
[Ulrich & Ulrich (1982), 'Little Brother', Genoveva Bol]

u   tz'-aj-Ø   b'in ich **k'aak'**.
3A put-TR.PFV-3B HSY in   fire
'She put it on **(the) fire**.'

Although the term 'fire' does not inherently identify a unique individual, in the context of a Mopan house where cooking has been mentioned, only one fire can be intended (see, e. g., Löbner 2011: 285).

3.3 Pragmatically (and also semantically) specific indefinites

Recall that in Dryer's hierarchy, semantically specific indefinites presuppose the existence of a referent (*I went to a movie last night*), as opposed to semantically nonspecific indefinites, which do not make this presupposition (*John is looking for a new house*). Semantically specific indefinites come in two pragmatic types. Pragmatically specific indefinites are those which will remain topical, i. e., are mentioned again in the discourse after they are introduced. Pragmatically non-specific indefinites are not mentioned again in the subsequent discourse.[18] By Dryer's definition (2014: e237), semantically nonspecific reference cannot be pragmatically specific.

3.3.1 ART alone for pragmatically specific indefinites

Although new referents that will remain topical are typically introduced with the *jun* + classifier construction (section 3.3.4), it is also possible for such a referent to be marked only with ART. This is illustrated in (21).

(21) Pragmatically specific new referent introduced with ART alone.
[Author's data, 'The Ring and the Fish', P. C.]

pues a   winik a   b'e=e   u chaan-t-aj-Ø
so   ART man   ART D.3.NV=EV 3A gaze-T-TR.PFV-3B

t-u   tzeel=e uy(prevocalic) il-aj-Ø=a   yan-Ø a   **b'ak=a**.
PREP-3A side=EV 3A   see-TR.PFV-3B=EV exist-3B ART bone=EV
'So the mentioned man looked next to him, he saw there were **[ART] bone[s]**.'

---

[18]We note that a category that is based on subsequent mention in the discourse is weighted toward connected discourse such as narrative.

Example (21) is from a story in which an ogre disguised as a woman has lured a man's brothers into the forest and killed them. The bones, mentioned here for the first time, are evidence that the brothers have been killed. As such they are extremely important to the storyline, and they are mentioned again as the story continues. Here the *jun* 'one' + classifier construction would be less felicitous, since more than one bone is involved (pluralization is optional in Mopan), but other numbers would be over-specific in this context.

3.3.2 Relativized deictic predicates for introducing an unfamiliar referent

Though rare for first mention, a relativized deictic can also occur in this context, as shown in (22).

(22) Relativized deictic for first mention of pragmatically specific referent.
    [Ventur (1976) 3:03, *A ayin etel aj Konejo* 'The Story of the Alligator and the Rabbit', E. S.]

pues **jun tuul**   **b'in a**   **winik a**   **b'e'=e,**
so   one CLF.ANIM HSY ART man   ART D.3.NV=EV

tan-Ø   b'in u   man-äl.
be.continuing-3B HSY 3A walk-INTR.IPFV

'So **this man**, he was wandering along.'

This is the first mention of the protagonist in a story. The effect of combining the numeral + classifier + ART construction, commonly used for first mention of a referent that will continue to be topical (section 3.3.4 below), with the relativized deictic *a b'e'* 'associated with neither speaker nor hearer and known through non-visual means', more often used for anaphoric reference, is similar to what Prince (1981) calls "indefinite *this*" in English.

3.3.3 Emphatic pronoun for pragmatically specific reference

The 3rd person emphatic pronoun *le'ek* may also appear at first mention of a referent. Example (23) occurs in the first line of the story, and the referent introduced with *le'ek* is presented as syntactically equivalent to the one which is introduced with the numeral + classifier construction, very frequently used for first mentions.

Eve Danziger & Ellen Contini-Morava

(23)   *Le'ek* for first mention.
        [Ventur (1976) 7:08, *A B'aalumoo'o* 'The Jaguars', A. T.]

        jum p'eel   k'in b'in,
        one CLF.INAN day HSY

        le'ek   a   b'aalum uy   et'ok
        3.EMPH ART jaguar   3A(prevocalic) companion

        jun tuul   aj   leon=o
        one CLF.ANIM GM.M lion=EV

        uy   ad'-aj-Ø-oo'   b'in ti   u   b'ajil
        3A(prevocalic) say-TR.PFV-3B-3.PL HSY PREP 3A self

        'One day **that which is (a) jaguar** together with a [NUM + CLF] lion,[19]
        they said to each other ...'

In this case, both the jaguar introduced with *le'ek* and the 'lion' (also afterwards called *b'aalum* 'jaguar') are highly salient, and are mentioned multiple times in the discourse that follows.

In light of (23) and other examples of first mention, *le'ek* 'be it/be the one' must therefore be understood as an indicator of emphasis rather than primarily one of definiteness.

3.3.4 Numeral classifier construction for pragmatically specific reference

The most common way in Mopan to introduce new referents that will remain topical in subsequent discourse is by means of the numeral *jun* 'one' (or other numeral where appropriate), followed by a numeral classifier and the nominal. The nominal may or may not also be preceded by ART. This is shown in (24).

---

[19]The word *leon* 'lion, jaguar' belongs to a subset of Mopan vocabulary that is lexically specified for gender. For such nouns a gender marker is essentially obligatory and has no relationship to definiteness (see Contini-Morava & Danziger 2018 for more on the Mopan gender markers).

(24) *Jun* 'one' + classifier with and without ART for pragmatically specific new referent.

   a. *Jun* 'one' + CLF with ART.
       [Ventur (1976) 5:09, *U Alka'b'eeb' Jesus*, 'The Chasing of Jesus', J. I.]

       pues k'och-ij   tub'a   yan-Ø
       so   arrive-3B.INTR.PFV where exist-3B

       jun teek   a   mäp=ä.
       one CLF.plant ART cocoyol_palm=EV

       'So he arrived at [a place] where there was **a [NUM + CLF + ART] cocoyol palm**.'

   b. *Jun* 'one' + CLF without ART.
       [Ventur (1976) 4:02, *U Kwentojil aj Konejo manyoso*, 'The Story of the Clever Rabbit', A. K'.]

       entonses b'in-ij   u ka'   käx-ä'.
       then   go-3B.INTR.PFV 3A again seek-3B.TR.IRR

       ke'en-Ø   yalam **jun teek   mäp.**
       be.located-3B under one CLF.plant cocoyol_palm

       'So he [puma] went off to look for him [rabbit] again. He [rabbit] was located under **a [NUM + CLF, no ART] cocoyol palm**.'

Even though each example introduces a (single) cocoyol palm that is referred to again in subsequent discourse, one includes ART and the other does not.[20] We propose a pragmatic explanation for presence/absence of ART in such cases. When a lexeme, due to its meaning, is likely to be construed as an entity, the article can be omitted if the referent has lower discourse salience than it would if it were marked by ART.[21]

We note, however, that lack of discourse salience in this sense does not correspond precisely to Dryer's 'pragmatic nonspecificity', since in both of the

---

[20]A referee asks whether *jun* 'one' in (24b) is perhaps a type of indefinite determiner rather than the numeral 'one'. Although the numeral + classifier construction illustrated here is often translatable with an indefinite article in English, this translation does not depend on presence vs. absence of ART. The cocoyol palms in (24a) and (24b) are both new discourse referents, and in neither case is their singularity being contrasted with other possible numbers. (Note also that other numerals can occur both with and without ART in a numeral + classifier construction in Mopan.)

[21]At the 2018 Workshop on Specificity, Definiteness and Article Systems Across Languages (40th Annual Meeting of the Deutsche Gesellschaft für Sprachwissenschaft) we provided some quantitative evidence in support of differential discourse salience of presence/absence of ART; see Contini-Morava & Danziger (forthcoming) for those data.

Eve Danziger & Ellen Contini-Morava

examples in (24), the referent is mentioned again in later discourse. Difference in discourse salience is, rather, a question of degree of importance of the referent as a protagonist in the discourse in question. To illustrate, in (24a) above, Jesus is fleeing from persecution and hides at the top of a cocoyol palm. In the ensuing narrative when the pursuers ask the tree what it is hiding, it responds in a misleading way so as to protect Jesus. The tree is a salient protagonist in the story. In (24b), the cocoyol palm never speaks or takes on animacy, and is eventually broken up and offered as food, losing its quality as an (individually identifiable) entity. The word *mäp* 'cocoyol palm' in this second case is determinerless when first mentioned, and —although it qualifies for Dryer's 'pragmatic specificity' because it is mentioned again in the same text— it is not an important character in the story. (Further indication of the difference in discourse salience between the trees in these examples is the fact that the first one is introduced as the main argument of its clause whereas the second is introduced in a prepositional phrase.)

3.3.5 Bare nominal for pragmatically specific reference

It is also possible for a bare nominal to introduce a new referent that will be mentioned again in the discourse, as shown in (25).

(25) Bare nominal for pragmatically specific referent.
[Ulrich & Ulrich (1982), 'Trip to Belize', José María Chowoj]

pwes ki' keen-oo' ti'i i jok-een toj ich **naj**.
well good 1.QUOT-3.PL 3.OBL and exit-1B already in house
' "Well, good!" I said to them and went out of (the) **house**.'

b'in-o'on pach **naj**.
go-1B.PL behind house
'We went behind (the) **house**.'

pues te'=i in jok-s-aj-oo' u foto
well D.LOC.3.NV=SCOPE 1A exit-CAUS-TR.PFV-3.PL 3A photo

ti k'och-ij a soldadoj=o.
PREP arrive-3B.INTR.PFV ART soldier=EV
'Well, I had taken their picture(s) when a policeman arrived.'

uch-ij u cha'an ich **naj**.
happen-3B.INTR.PFV 3A gaze in house
'He looked around in (the) **house**.'

In (25), the narrator describes a visit to an acquaintance, whose house is an example of a referent whose specificity and uniqueness are given by the context. The house is introduced with the bare nominal *naj*, and is mentioned two more times again with a bare nominal. Despite its specificity, and despite the fact that it is mentioned more than once, the house is not an important participant in this narrative: it is mentioned merely in its capacity as location.[22]

3.4 Semantically specific but pragmatically nonspecific referents

Recall once again that for Dryer (2014), a semantically specific referent involves an entailment of existence but that such referents can be either pragmatically specific (recurs in discourse —Mopan examples of such cases were treated in the previous section), or pragmatically nonspecific. A semantically specific but pragmatically nonspecific referent in Dryer's terminology does not remain topical in the discourse: it is never mentioned again. In narrative at least, referents that receive only one mention are unlikely to be important protagonists in the discourse in which they occur. Dryer's pragmatic nonspecificity therefore coincides to a great extent with our 'low discourse salience' (although we have seen that the converse is not the case: Dryer's pragmatic specificity can cover instances both of low and of high discourse salience, see examples (24a-b)).

3.4.1 ART for semantically specific, pragmatically nonspecific referents

ART by itself is rarely used for referents which will not be mentioned again in the discourse (pragmatically nonspecific indefinites). This is not surprising, given that such expressions are by definition not salient in the ensuing discourse and given that use of ART correlates with discourse salience. While not common, it is nevertheless possible for ART to appear with a semantically specific NP that is not mentioned again, as in example (26).

(26)   ART for semantically specific, pragmatically nonspecific referent.
[Ventur (1976) 2:04, *U Kwentojil ix Pulya'aj* 'The Story of the Witch', A. K'.]

tan-Ø   b'in u tzäj-ik-Ø   a   ja'as=a.
be.continuing-3B HSY 3A fry-TR.IPFV-3B ART banana=EV

tan-Ø   b'in u tzäj-ik-Ø   **a   kamut=u.**
be.continuing-3B HSY 3A fry-TR.IPFV-3B ART sweet.potato=EV

'She was frying plantain(s). She was frying **sweet potato(es)**.'

---

[22]Note that the preposition *ich* 'inside' may also occur with an ART-marked nominal, i. e., it is not obligatorily followed by a bare nominal.

Example (26) is from a story in which two children, abandoned by their father in the forest, come upon a house where an old blind woman is cooking plantain and sweet potato (recall that plural specification is optional in Mopan). They eventually steal the plantains but sweet potato is not mentioned again in the story.

3.4.2 Relativized deictic for semantically specific, pragmatically nonspecific indefinites

Though rare, it is also possible for the relativized deictic *a b'e'* to introduce a new discourse referent that is not mentioned again, as shown in example (27).

(27)   Relativized deictic for pragmatically nonspecific referent.
[Ventur (1976) 1:03, *Jun tuul Winik etel Ma'ax* 'A Man and a Monkey', E. S.]

entonses ti   ka'   b'in jun tuul   ilik b'in
then   PREP again HSY one CLF.ANIM just HSY

a   koch   p'at-al   ti   uk'-ul
ART last.one abandon-INTR.PFV PREP drink-INTR.IPFV

ichil **a   tunich a   b'e'=e**
inside ART rock   ART D.3_NV=EV

entonses pues te'=ij=i,
then   so   LOC.3.NV=SCOPE=EV

b'in-ij   b'in u chiit-t-ej
go-3B.INTR.PFV HSY 3A speak.to-T-3B.TR.IRR

a   b'e'   a   ma'ax=a.
ART D.3.NV art monkey=EV

'Then when there was just one last (monkey) left behind, drinking from **that rock**, then at that point, he (hero) went and spoke to that monkey.'

The hero of this story, a hunter who is thirsty, has spotted some monkeys drinking at a location that is not specified. Fearing the monkeys, he hides until most of them depart, leaving one behind. In (27), the narrator mentions for the first time a rock that the monkey was drinking from. Although the hero eventually befriends the monkey, the rock is not mentioned again. This would not be an example of uniqueness being given by the context, like the household fire in example (20), because drinking at a stream in the forest does not presuppose drinking from a rock. Possibly the demonstrative in (27) is meant to suggest that this monkey is in the same place where the others had been, even though that place was not explicitly described.

3.4.3 Emphatic pronoun for semantically specific, pragmatically nonspecific indefinites

The independent pronoun *le'ek* can be used for semantically specific but pragmatically nonspecific indefinites. In example (28), a trickster rabbit convinces a puma that a large rock is in danger of falling over. But in fact the rock is firm —a cloud passing overhead has created the illusion of instability.

(28) *Le'ek* for semantically specific, pragmatically nonspecific indefinite.
  [Ventur (1976) 6:01, *Aj Koj etel aj konejo* 'The Puma and the Rabbit', M. X.]

  pero le'ek   a   muyal
  DISC 3.EMPH ART cloud

  a   tan-Ø   u b'eel   ti   u wich=i.
  ART be.continuing-3B 3A go.IPFV PREP 3A face=EV

  'But it was **that which is a cloud** that was passing over its face.'

This sentence constitutes the first and only mention of the cloud. Here emphasis on its identity (as a cloud) contrasts with the appearance of instability of the rock. Although the cloud is not mentioned again, use of the emphatic pronoun highlights its role in the trick being played on the puma.

3.4.4 Bare nominal for pragmatically nonspecific referents

Pragmatically nonspecific referents are typically referred to with bare nominals in Mopan, as shown in example (29).

(29) Bare nominal for semantically specific but pragmatically nonspecific referent.
  [Ventur (1976) 1:01, *Aj Jook'* 'The Fisherman', R. K'.]

  pues jak'-s-ab'-ij
  then frighten-CAUS-PASS-3B.INTR.PFV

  b'in uy(prevocalic) ool   u_men **kan**.
  HSY 3A   feeling by   snake

  'Then he was startled by **(a) snake**.'

The snake referred to here is never mentioned again in the story. It does not contrast with any previously established expectation (as the cloud does in example (28)), nor does it play an important role in the plot.

We have already noted that a notion of discourse salience —importance of the referent as a participant in the surrounding narrative— governs the distribution of bare nominals and of ART in Mopan narratives, and that this is not

necessarily coterminous with Dryer's contrast between pragmatic specificity and pragmatic nonspecificity (whether a referent is or is not mentioned again in subsequent discourse). The examples in this and the previous section show once again that repeated mention, or lack of it, is at best an indirect marker of discourse salience: a referent may be mentioned only once but play a significant role (example (28), the cloud), and a referent may be mentioned more than once but play a peripheral role (example (25), the house).

3.5 Semantically nonspecific indefinites

Semantically nonspecific indefinites make no claim as to the actual existence of the referent.

3.5.1 ART for semantically nonspecific reference

Even though the category of semantically nonspecific is at the least definite end of Dryer's reference hierarchy, it is possible for ART to occur in this context in Mopan, as shown in example (30).

(30) Use of ART for semantically nonspecific indefinite.
      [Verbeeck (1999: 11), *U Kwentajil a Santo K'in* 'The Story of the Holy Sun',
      narrated by Alejandro Chiac]

      sansamal tatz' tan-Ø   b'in u   b'el
      daily   far   be.continuing-3B HSY 3A go.IPFV

      u   tz'on-o'   a   **yuk=u**.
      3A shoot-3.BTR.IRR ART antelope=EV

      'Every day he went far [into the woods] to shoot **an [ART] antelope**.'

      In (30), there is no entailment of existence of an antelope. The fact that no particular antelope is being referred to (despite use of ART) can be inferred from the imperfective marking on the action of hunting, along with the time reference 'every day', which make it highly unlikely that the same antelope would be involved on each occasion of hunting.

3.5.2 Relativized deictic for semantically nonspecific reference

Given the strong association of *a b'e'* 'deictic stative 3rd person known through other than visual means' with anaphoric reference (section 3.1.2 above), and its highlighting effect elsewhere, we would not expect it to be used for nonspecific

referents, and indeed we did not find any examples of *a b'e'* used for this purpose in our data.

3.5.3 Independent pronoun *le'ek*

We have found no examples of the independent pronoun *le'ek* being used for semantically nonspecific referents. The semantics of this form ('that which is 3rd person') perhaps categorically preclude such usage.

3.5.4 Numeral classifier construction

It is possible, though rare, for a numeral + classifier construction to be found with semantically nonspecific referents, as shown in (31).

(31) Numeral + classifier construction for semantically nonspecific indefinite. [Ventur (1976) 1:05, *U kwentojil Santo K'in y Santo Uj* 'The Story of the Holy Sun and the Holy Moon', R. K.']

   in tat=a,   u k'ati **jun tuul**   **ix**   **ch'up=u**.
   1A father=EV, 3A want one CLF.ANIM GM.F woman=EV
   'My father, he wants **a woman/wife**.'

   Example (31) is uttered by a vulture to a woman whom he hopes to persuade to marry his father. The father does not know this woman, so the woman referred to here is nonspecific.

3.5.5 Bare nominal for semantically nonspecific indefinite

In our discussion of examples (24a-b) above, we mentioned that use of ART vs. a bare nominal correlates with discourse salience in the case of semantically specific referents. This contrast also applies to semantically nonspecific referents, as shown in example (32). In (32), the speaker lists several hypothetical animals that he wants to hunt. Some are marked by ART and some are bare nominals.

(32)   Semantically nonspecific indefinite reference.
[Author's data, *Ix Che'il etel Bäk'* 'Wild Woman', J. S.]

ix   kolool,   **k'änb'ul**, **kox**
GM.F partridge pheasant cojolito (type of game bird)

etel **a**   **kek'enche'** etel **a**   **yuk=u**
with ART wild.pig   with ART antelope=EV

le'ek   kuchi in k'ati tz'on-oo' pere ma' yan-Ø   kut'an.
3.EMPH DISC 1A want shoot-3.PL but NEG exist-3B 3.QUOT

' "[GM] Partridge[23], [no ART] **pheasant**, [no ART] **cojolito** [type of game
bird], and [ART] **wild pig**, and [ART] **antelope**, those are what I really
want to hunt, but they aren't there!" he said.'

Even a hypothetical or non-existent referent can figure more or less centrally in discourse. Recall that in example (30) (section 3.5.1 above), the protagonist repeatedly hunts for an antelope because he wants to impress a young woman with his prowess as a hunter. Even though no specific antelope is being referred to in that example, a hypothetical antelope is important to the plot: the protagonist eventually tries to trick the woman by carrying a stuffed antelope skin past her house. In (32), all the animal terms have equal status from the point of view of nonspecificity and all play the same syntactic role, but the referents differ in discourse salience: ART is omitted before the names of the birds and retained before the names of the larger mammals that are more desirable as game. The wild pig and antelope are also treated differently from the birds in that they are each introduced with the conjunction *etel* 'and/with'.

4 Summary and conclusions

Table 1 is a summary, according to Dryer's hierarchy, of the distribution of expressions that contribute to referential anchoring in Mopan narratives, as discussed in this chapter. The table is not intended to make comprehensive quantitative claims. The double pluses mean that certain types of examples are easily found via investigation of multiple Mopan texts; the single pluses require more diligent searching. Minus signs mean that we have not found any such examples despite diligent searching.

---

[23]The word *kolool* 'partridge' is a feminine noun. Recall (supra note 19) that for this subset of nouns a gender marker is essentially obligatory.

4 Referential anchoring without a definite article: The case of Mopan (Mayan)

Table 1: Summary of referential anchoring in Mopan in relation to Dryer's (2014) reference hierarchy (with subdivisions for relative discourse prominence added).

	Anaphoric definites	Non-anaphoric definites	Semantically and pragmatically specific indefinites	Semantically specific, pragmatically nonspecific indefinites	Semantically nonspecific indefinites
ART	++	++	+[a]	+	+
Relativized deictic *b'e'*	++	+	+	+	−
Emphatic pronoun *le'ek*	++	+	+	+	−
NUM + CLF + ART	−	−	+	−	−
NUM + CLF	−	−	+	+	+
Bare nominal	+	+	+	+	+

[a]Most likely to occur with high discourse salience.

The forms that have the most consistent connection to messages of definiteness are the relativized deictic predicate *b'e'* and the emphatic pronoun *le'ek*, found primarily at the most definite end of the hierarchy (and not found at the least definite end), and the *jun* 'one' + classifier construction, which is found primarily at the less definite end of the hierarchy, with a preference for contexts of specificity.

Both ART and the bare nominal option appear all across the hierarchy, from the maximally identifiable end (highly predictable anaphoric definites) to the least identifiable end (semantically and pragmatically nonspecific). Their distribution is not compatible with a semantics of (in)definiteness.[24] Instead, ART is required in order to entitize lexical content that would not otherwise be construed as an entity/argument (see section 2.1, and Contini-Morava & Danziger forthcoming). With lexical content that lends itself to construal as an entity, ART is optional, and we have argued that its presence/absence is sensitive to the discourse salience of the referent. ART's tendency to occur most often at the definite end of Dryer's hierarchy follows from the fact that entities that are part of common ground between speaker and addressee tend also to be relatively salient in discourse. But local contexts or nonlinguistic knowledge can lend salience even to nonspecific indefinites. ART, in short, is not a form that is dedicated to signaling contrasts on the definiteness dimension. Nevertheless, ART alone (without

---

[24]For further discussion of Mopan in this connection, see Contini-Morava & Danziger (forthcoming).

Eve Danziger & Ellen Contini-Morava

relativized deictic or numeral classifier construction) is one of the most common constructions with which arguments occur in Mopan.

Meanwhile, the fact that the bare nominal construction is also allowable across all of the positions in the hierarchy makes clear that although dedicated means for indicating definiteness or indefiniteness exist in Mopan, they are always optional. We conclude, then, that the status of a discourse referent with regard to relative familiarity, referentiality, specificity and related notions normally considered as aspects of 'definiteness' may be left unspecified in Mopan. If it is found necessary to make such determination in a given case, this must frequently be accomplished through pragmatic inference, rather than via information explicitly signaled by particular grammatical forms.

Acknowledgments

Abbreviations and glosses

A = set A case-role marker (Actor or Possessor)

ANIM = animate

B = set B case-role marker (Undergoer)

D = deictic

DISC = discourse particle

EMPH = emphatic pronoun

EV = prosodic echo vowel

GM = gender marker

HSY = hearsay particle

INAN = inanimate

INCH = inchoative

INT = intensifier

NV = non visible

POSS = Possessive

PREP = general preposition

SCOPE = scope clitic

T = transitivizer

V = visible

References

Alexiadou, Artemis. 2005. Possessors and (in)definiteness. *Lingua* 115(6). 787–819.

Bricker, Victoria Reifler. 1981. The source of the ergative split in Yucatec Maya. *Journal of Mayan Linguistics* 2. 83–117.

Contini-Morava, Ellen & Eve Danziger. Forthcoming. Determination, bare arguments, and discourse salience in Mopan Maya.

Contini-Morava, Ellen & Eve Danziger. 2018. Non-canonical gender in Mopan Maya. In Sebastian Fedden & Jenny Audring (eds.), *Non-canonical gender systems*, 129–146. Oxford: Oxford University Press.

Danziger, Eve. 1994. Out of sight, out of mind: Person, perception and function in Mopan Maya spatial deixis. *Linguistics* 32(4/5). 885–907.

Danziger, Eve. 1996. Split intransitivity and active-inactive patterning in Mopan Maya. *International Journal of American Linguistics* 62(4). 379–414.

Danziger, Eve. 2008. A person, a place or a thing? Whorfian consequences of syntactic bootstrapping in Mopan Maya. In Melissa Bowerman & Penelope Brown (eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability*, 29–48. Mahwah, NJ: Lawrence Erlbaum Associates.

Davis, Henry, Carrie Gillon & Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4). e180–e226.

Dryer, Matthew S. 2007. Noun phrase structure. In Timothy Shopen (ed.), *Language typology and syntactic description, Vol. II: Complex Constructions*, 2nd Edition, 151–205. Cambridge: Cambridge University Press.

Dryer, Matthew S. 2014. Competing methods for uncovering linguistic diversity: The case of definite and indefinite articles (commentary on Davis, Gillon, and Matthewson). *Language* 90(4). e232–e249.

England, Nora C. & Stephen R. Elliott. 1990. Lecturas sobre la lingüística Maya. Centro de Investigaciones Regionales de Mesoamérica, Guatemala.

Gillon, Carrie. 2009. The semantic core of determiners. In Jila Ghomeshi, Ileana Paul & Martina Wiltschko (eds.), *Determiners: Universals and variation*, 178–213. Amsterdam: John Benjamins.

Gillon, Carrie. 2013. *The semantics of determiners: Domain restriction in Sḵwx̱wú7mesh*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Givón, Talmy. 1978. Definiteness and referentiality. In Joseph Greenberg, Charles Ferguson & Edith Moravcsik (eds.), *Universals of human language*, vol. 4: Syntax, 291–330. Stanford, CA: Stanford University Press.

Gundel, Jeanette, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Eve Danziger & Ellen Contini-Morava

Hanks, William. 1990. *Referential practice: Language and lived space among the Maya.* Chicago: University of Chicago Press.

Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction.* London: Croom Helm.

Himmelmann, Nikolaus. 2008. Articles. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals / Sprachtypologie und Sprachliche Universalien*, 831–841. Berlin: De Gruyter.

Hofling, Charles A. 2006. Juan Ceniza, a Mopan Maya story. *Latin American Indian Literatures Journal* 22(1). 22–42.

Jelinek, Eloise. 1995. Quantification in Straits Salish. In Emmon Bach, Eloise Jelinek, Angelika Kratzer & Barbara H. Partee (eds.), *Quantification in natural languages*, vol. 2, 487–540. Dordrecht: Kluwer.

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

Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.

Lois, Ximena & Valentina Vapnarsky. 2006. Root indeterminacy and polyvalence in Yukatekan Mayan languages. In Ximena Lois & Valentina Vapnarsky (eds.), *Lexical categories and root classes in Amerindian languages*, 69–116. Berlin: Peter Lang.

Lucy, John A. 1994. The role of semantic value in lexical comparison: Motion and position roots in Yucatec Maya. *Linguistics* 32. 623–656.

Lyon, John. 2015. Okanagan determiner phrases and domain restriction. *International Journal of American Linguistics* 81(2). 187–219.

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

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

Prince, Ellen. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.

Sasse, Hans-Jürgen. 1988. Der irokesische Sprachtyp. *Zeitschrift für Sprachwissenschaft* 7(2). 173–213.

Shaw, Mary. 1971. Texts in Mopan. In Mary Shaw (ed.), *According to our ancestors: Folk texts from Guatemala and Honduras*, 410–430. Norman, OK: Summer Institute of Linguistics.

Ulrich, Matthew & Rosemary Ulrich. 1982. *Stories of the sacred, serious, sensational and silly from Mopan Maya texts.* Verano, Guatemala: Instituto Lingüístico de Verano en Centroamérica.

Ulrich, Matthew, Rosemary Ulrich & Charles Peck. 1986. *Mopan Mayan verbs.* Guatemala City: Summer Institute of Linguistics.

Ventur, Pierre. 1976. *Mopan Maya folklore texts.* Washington D.C.: Smithsonian Institution. (Accession #80-38).

Verbeeck, Lieve. 1999. *U kwentajil a Santo K'in: the story of the Holy Sun.* Waddinxveen NL: Mopan Maya Oral Literature. Narrated by Alejandro Chiac, transcribed and translated by Lieve Verbeeck with the assistance of Valeria Sho.

Chapter 5

The specificity marker *-e* with indefinite noun phrases in Modern Colloquial Persian

Klaus von Heusinger
Universität zu Köln

Roya Sadeghpoor
Universität zu Köln

Persian has two indefinite markers, the prenominal *ye(k)* and the suffixed *-i*. Both forms express particular kinds of indefiniteness, as does their combination: for Modern Colloquial Persian, indefinites ending in *-i* express a non-uniqueness or anti-definite implication and behave similarly to *any* in English. *Ye(k)*, on the other hand, expresses an at-issue existence implication and behaves similarly to the English *a(n)* (Jasbi 2016). The combination of *ye(k)* and *-i* expresses an ignorance implication. Modern Colloquial Persian has the specificity marker *-e*, which can be combined with *ye(k) NP*, as well as with the combined form of *ye(k) NP-i*, but not with (solitary) *NP-i* (Windfuhr 1979; Ghomeshi 2003). In this paper, we investigate the function of the indefinite form when combined with the specificity marker *-e*, namely *ye(k) NP-e* and *ye(k) NP-e-i*. We present two pilot studies that tested our hypothesis, which is that the contrast between these two specific forms depends on whether the specificity is speaker-anchored, as for *ye(k) NP-e*, or non-speaker anchored, as for *ye(k) NP-e-i*. The results of the two studies provide weak support for this hypothesis, and provide additional evidence for the fine-grained structure of specificity as referential anchoring (von Heusinger 2002).

1 Introduction

Persian is a language with no definite marker and two indefinite markers. In Modern Colloquial Persian, the prenominal indefinite article *ye(k)* 'a(n)' marks

an NP as indefinite and expresses an existential entailment 'there is at least one N', as in (1), similar to a noun phrase with the indefinite article in English. In Modern Colloquial Persian, the suffixed (or enclitic) marker *-i* is interpreted as a negative polarity item (NPI), as in (2), similar to the English *any* (Jasbi 2014; Lyons 1999; Windfuhr 1979). Both indefinite markers can be combined into a complex indefinite, consisting of *ye(k) NP-i*, which is interpreted as a free-choice item, as in (3), or with a certain 'flavor' of referential ignorance, as in (4), similar to *some or other* in English (Jasbi 2016).[1]

(1) Emruz yek māšin tu xiābun didam. *ye(k) NP*
today a car at street saw.1SG (existential)
'Today I saw a car on the street.'

(2) Māšin-i ro emruz tu xiābun didi? *NP-i*
car-i rā today at street saw.2SG (negative polarity item)
'Did you see any cars on the street today?'

(3) Context: Ali wants to play the lottery. Reza is explaining to him how it is played.

Ye šomāreh-i ro entexāb kon va injā alāmat bezan. *ye(k) NP-i*
ye number-i rā choose do.2SG and here mark do.2SG (free choice)
'Choose a number and mark it here.'

(4) Yek bače-i tu xiābun gom šode bud. *ye(k) NP-i*
ye child-i at street lost became.3SG was.3SG (ref. ignorance)
'A/some child was lost in the street.'

Modern Colloquial Persian has the optional suffix *-e*, which we take to express specificity. The literature assumes different functions of this suffix, such as a demonstrative, a definite, or a referential function (Windfuhr 1979: 40; Hincha 1961: 173-177; Lazard 1957: 163; Ghomeshi 2003: 67) or familiarity of the referent (Hedberg et al. 2009) as in the anaphoric noun *pesar-e* in (5):

(5) Emruz ye pesar va ye doxtar ro did-am. Pesar-(e) tās bud.
today a boy and a girl rā saw-1SG boy-e bald was.3SG
'Today I saw a boy and a girl. The boy was bald.'

---

[1]Persian has a differential object marker *-ra/-ro/-a/-o* (generally glossed as *-rā* or as OM, DOM or ACC), which is obligatory with definite and specific direct objects, and optional with non-specific indefinite direct objects (Ghomeshi 2003; Karimi 2003; 2018; Lazard 1957; 1992; Windfuhr 1979).

The suffix *-e* is typically used with demonstrative and definite noun phrases, but it can also be combined with the indefinite constructions discussed above, which we take as evidence that it expresses specificity (or referential indefiniteness): (i) its combination with the indefinite marker *ye(k)*, i. e., *ye(k) NP-e*, as in (6), yields a specific reading; (ii) it cannot be combined with suffixed indefinite *-i*: **NP-e-i*, as in (7), due to the incompatibility of the specific function of *-e* and the free-choice function of *-i*; (iii) the specific marker *-e*, however, can be combined with the complex indefinite *ye(k) NP-e-i*, yielding a specific reading in (8), which is very similar to (6).

(6) **Ye pesar-e** injā kār mikone.
    a boy-e here work do.3SG
    'A specific boy works here.'

(7) Diruz   **māšin-i** jole   dare xune didi?
    yesterday **car-i** front.of door.of home saw.2SG

    – *Na, man hič māšin-e-i nadidam.
       no I **any car-e-i** not.saw.1SG
    'Did you see any cars in front of the house door yesterday?'
    Intended reading: 'No, I didn't see any specific car.'

(8) Emruz **ye māšin-e-i** az   post   behem zad.
    today **a car-e-i** from behind to.me collided.3SG
    'Today a specific car collided into me from behind.'

This data, then, raises the following questions. First, what are the differences in the meanings of the three forms expressing indefiniteness in (1) through (4) in Modern Colloquial Persian? Second, what is the contribution of the marker *-e*? Does it express specificity or a different semantic pragmatic notion, such as referentiality, demonstrativeness, topicality, or partitivity? Third, what is the function of the marker *-e* with indefinite constructions, and, more specifically, what is the difference between the two (specific) indefinite constructions *ye(k) NP-e* and *ye(k) NP-e-i*? We assume the following functions of the three indefinite constructions (cf. Jasbi 2014; Lyons 1999; Windfuhr 1979): (i) the indefinite marker *ye(k)* signals a regular indefinite, i. e., it expresses an existential entailment, but does not encode specificity (like the English *a(n)*); (ii) the suffixed marker *-i* is a negative polarity item (like the English *any*); (iii) the combination of the two markers, resulting in *ye(k) NP-i*, shows an ignorance or free-choice implicature.

Second, we assume that the marker *-e* in Modern Colloquial Persian signals specificity in terms of "referential anchoring", in accordance with von Heusinger

Klaus von Heusinger & Roya Sadeghpoor

(2002). An indefinite is referentially anchored if the speaker, or another prominent discourse referent, can readily identify the referent. This more fine-grained notion of specificity allows us to formulate our Hypothesis 1, about the semantic difference between the two indefinite constructions with the specificity marker, namely *ye(k) NP-e* and *ye(k) NP-e-i*; the specific indefinite construction *ye(k) NP-e* only reflects the intention of the speaker (or speaker-oriented specificity), while the form *ye(k) NP-e-i* only expresses the intention of another salient discourse participant (i. e., non-speaker-oriented specificity).

In section 2, we provide a brief overview of the variety of indefinites found in different languages, as well as the ranges of different functions that indefinites can take. In particular, we focus on the contrast between speaker-oriented specificity and non-speaker-oriented specificity. In section 3, we discuss the different functions of the indefinite markers in Modern Colloquial Persian and modify the approach of Jasbi (2016). In section 4, we present some relevant data for the use of the marker *-e* in Modern Colloquial Persian, and in section 5, we present the two pilot studies that addressed our hypotheses about the speaker-oriented specificity of these forms. Finally, section 6 provides a discussion and a conclusion.

2 Indefinites in the languages of the world

2.1 Indefinite articles

Languages differ as to whether or not they mark indefinite noun phrases with special morphological means, such as indefinite articles. In Dryer's (2005) WALS sample, 57% of the languages do not have indefinite articles.

Table 1: Types of article systems (Dryer 2005)

Type of article system	Instances	Percentages
Indefinite word distinct from numeral for 'one'	91	19%
Numeral for 'one' is used as indefinite article	90	19%
Indefinite affix on noun	23	5%
No indefinite article but definite article	81	17%
Neither indefinite nor definite marker	188	40%
Total	473	100%

Among the 43% of languages that do have an indefinite marker, we find some that have more than one indefinite marker or article, which often expresses the contrast between a specific reading, as in (9a), and a non-specific reading, as in (9b), from Lakhota, North America (Latrouite & Van Valin 2014: 405).[2]

(9) a. Wówapi waŋ   o⟨Ø-wá⟩le.   [olé 'look for']
     book   a[+specific] look.for⟨INAN-1SG.A⟩
     'I'm looking for a [particular] book.'

    b. Wówapi waŋží   o⟨Ø-wá⟩le.
     book   a[−specific] look.for⟨INAN-1SG.A⟩
     'I'm looking for a book [any book will do].'

Moroccan Arabic provides a three-way system of indefinite marking: (i) bare nouns are not marked for specificity, as in (10a); (ii) a specific indefinite article *wahed-l*, composed of the numeral 'one' and the definite article, as in (10b); (iii) a non-specific indefinite article *shi*, derived from the word for 'thing', as in (10c) (from Fassi-Fehri 2006; see Brustad 2000: 26-31 for other Arabic dialects):

(10) 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
     Maryam wanted-F to-F-marry with-**one the-man** but

     ma lqa-t-u-sh.
     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
     Maryam wanted-F to-F-marry with-**some man** but

     ma lqa-t-u-sh.
     not met-her-him-NEG
     'Maryam wanted to marry a (non-spec.) man but she hasn't found
     one/(*him).'

---

[2]Abbreviations: A 'actor', INAN 'inanimate'.

We will argue in this paper that Modern Colloquial Persian not only exhibits the specific vs. non-specific contrast, as in Lakhota and Moroccan, but also allows us to morphologically mark a more fine-grained structure of specificity, namely whether the specific indefinite is oriented to the speaker or to some other prominent discourse referent within the context.

2.2 Speaker vs. non-speaker-oriented specificity

German, like English and other languages, has just one indefinite article (11a). However, it has other means of marking the specificity or referentiality of an associated noun phrase. While the regular indefinite in (11a) allows for both a wide- and a narrow-scope reading of the indefinite, the indefinite demonstrative in (11b) clearly signals a referential reading and forces a wide-scope reading:

(11) a. Jeder Student sagte ein Gedicht von Pindar auf.
          'Every student recited a poem by Pindar.'
      b. Jeder Student sagte dieses Gedicht von Pindar auf.
          'Every student recited this$_{indef}$ poem by Pindar.'

Many languages also have special adjectives that can induce different degrees of specificity. Ebert et al. (2013: 31) discuss the differences between the German adjectives *ein bestimmter* and *ein gewisser*, both of which the authors translate as 'a certain', even though the English translation does not reflect the subtle differences in meaning of the German adjectives. Their main observation is that both adjectives force the indefinite noun phrase to scope over the intentional verb *suchen* 'search' (12a-b), while the regular indefinite also allows for the narrow-scope reading, as in (12c):

(12) a. Peter sucht eine bestimmte CD / zwei bestimmte CDs / bestimmte CDs.
          Peter searches a BESTIMMT CD / two BESTIMMT CDs / BESTIMMT CDs.
          'Peter is looking for a certain CD / two certain CDs / certain CDs.'

                                                              ∃ > SEARCH
      b. Peter sucht eine gewisse CD / zwei gewisse CDs / gewisse CDs.
          Peter searches a GEWISS CD / two GEWISS CDs / GEWISS CDs.
          'Peter is looking for a certain CD / two certain CDs / certain CDs.'

                                                              ∃ > SEARCH
      c. Peter sucht eine CD / zwei CDs / CDs.
          Peter searches a CD / two CDs / CDs.
          'Peter is looking for a CD / two CDs / CDs.'   SEARCH > ∃, ∃ > SEARCH

The authors claim that the main difference between *ein bestimmter* and *ein gewisser* has to do with the bearer of the referential intention of that indefinite. For *ein gewisser*, only the speaker of the sentence can have that referential intention. For *ein bestimmter*, in contrast, the speaker or some other salient discourse agent, such as the subject of the sentence, can have this intention. This can be shown by the incompatibility of *ein gewisser* with speaker ignorance in (13b). The most natural reading of (13a) is that Peter knows which CD, but the speaker does not. So, the speaker only reports the assertion that there is some source (e. g., the subject) that has this referential intention.

(13) a. Peter sucht eine bestimmte CD, aber ich weiß nicht, welche.
          'Peter is looking for a BESTIMMT CD, but I do not know which one.'

      b. Peter sucht eine gewisse CD, #aber ich weiß nicht, welche.
          'Peter is looking for a GEWISS CD, #but I do not know which one.'

We can rephrase Ebert et al.'s observation in terms of "referential anchoring" in von Heusinger (2002; 2011; see also Onea & Geist 2011). The idea is that specific indefinites are anchored to the discourse referent that holds the referential intention about the identity of the referent. In a default case, indefinites are anchored to the speaker of the utterance. However, they can also be anchored to some other salient discourse referent, such as the subject of the sentence or other (implicit) referents. (For more on the notion of salience or prominence in discourse, see von Heusinger & Schumacher (2019).) We use this notion of speaker-oriented specificity vs. non-speaker-oriented specificity to account for the differences between the two specific indefinite constructions in Modern Colloquial Persian. That is, we will draw parallels between the two specific indefinite constructions in Modern Colloquial Persian and the contrast found for the German specificity adjectives *ein gewisser* vs. *ein bestimmter*.

3 Types of indefinites in Persian

Persian is a language with two dominant registers, spoken and written Persian, both of which have informal and formal forms that are very distinct (Jasbi 2014; Lazard 1957; 1992; Modarresi 2018; Nikravan 2014; Windfuhr 1979). The language that we investigate in this paper is Standard Modern Colloquial Persian. The function of the indefinite marker varies with register; the specificity marker *-e* is only used in Modern Colloquial Persian. In this section, we provide a brief overview of the way definiteness is expressed, the different indefinite forms in

Klaus von Heusinger & Roya Sadeghpoor

Standard Written Persian, and the use and function of indefinite forms in Modern Colloquial Persian.

3.1 Definiteness in Persian

Persian does not have a definite article, but it has two markers for indefiniteness (see the next section). To express definiteness, then, Persian typically uses bare noun phrases. This holds for different kinds of definite noun phrases. The definite in (14a) is a familiar definite, (14b) is a typical bridging definite, (14c) shows a unique definite, and (14d) is an example of generic use.

(14)   a.   Anne yek xune xarid. **Xune** labe marze kešvare.
          Anne a   house bought **house** on   edge.of country.be.3sɢ
          'Anne bought a house. The house is at the border of the country.'

     b.   Anne rafte   bud   ye marāseme arusi.   **Arus** xeyli
          Anne went.3sɢ ᴀᴜx.3sɢ a   ceremony marriage **bride** very

          xošgel   bud.
          beautiful was.3sɢ
          'Anne went to a marriage ceremony. The bride was very beautiful.'

     c.   **Māh**   xeyli rošan mideraxše.
          moon very   bright ᴘʀoɢ.shine.3sɢ
          'The moon shines very brightly.'

     d.   **Dianāsor** 60 milion sāle qabl monqarez šode.
          dinosaur 60 million year ago extinct   became.3sɢ
          'Dinosaurs became extinct 60 million years ago.'

There is controversy among scholars as to whether, in Persian, bare nouns are inherently definite (Krifka & Modarresi 2016), or underspecified with respect to definiteness and genericity (Ghomeshi 2003). Although it is not clear whether or not the non-specific indefinite nature of bare nouns can be detached from their generic (kind) reading, Dayal (2017) argues, using Hindi as an example, against the view that bare nouns are ambiguous and can have either a definite or an indefinite reading. She concludes that bare singulars in articleless languages like Hindi are definite and not indefinite (specific/non-specific), and that their apparent indefiniteness is construction-specific or restricted to bare plurals. Šimík & Burianová (2018) claim that in Czech, bare NPs, where they are indefinite, cannot be specific. Rather, bare NPs are either definite or indefinite non-specific, which is in line with Dayal's argument. Šimík & Burianová (2018), finally, annotate bare nouns

for (in)definiteness, and their findings suggest that the definiteness of a bare noun is affected by its absolute position in the clause, and that indefinite bare NPs are unlikely to occur in clause-initial position (see also Borik et al. 2020 [this volume]). Note that this is also applicable to Persian: Persian bare nouns can express a non-definite reading, as in (15a) with a kind-reading of 'book', or a definite-reading of 'book' as in (15b). Note that a bare noun in the preverbal direct object position is typically interpreted as pseudo-incorporated (in the sense of Massam 2001) as in (15c), while a definite reading must be signaled by the object marker *-rā* as in (15d) (see Modarresi 2014 for an analysis of bare direct objects in Persian):

(15) a. Roo miz   ketābe.
          on   table book.be.3SG
          'There is a/some book on the table.'

      b. Ketāb roo mize.
          book on table.be.3SG
          'The book is on the table.'

      c. Ali ketāb xarid.
          Ali book bought.3SG
          'Ali bought book/books.'

      d. Ali ketāb-rā xarid.
          Ali book-rā bought.3SG
          'Ali bought the book.'

3.2 Indefiniteness in Standard Written Persian

Standard Written Persian has the suffixed[3] indefinite marker *-i*, which has quite a large range of functions, and the independent lexeme *ye(k)*, which derives from the numeral *yek*, but behaves like a regular indefinite article.[4] Both forms can be

---

[3]There is some controversy as to whether *-i* is suffixed or enclitic. Herein we follow the works of Ghomeshi (2003); Hincha (1961); Karimi (2003); Paul (2008). This does not affect our analysis in any way.

[4]Ghomeshi (2003: 64–65) shows that the indefinite article *ye(k)* is different from the numeral *yek*. The former can appear without a classifier (i), which is obligatory for numerals, as in (ii) (see also Bisang & Ngoc Quang 2020 [this volume] for Vietnamese), and the indefinite article can also appear with plurals, as in (iii).

   (i) ye-(ta) ketab   (ii) se-*(ta)   ketab   (iii) ye ketab-ha-i
        a-(CL) book   three-*(CL) book   a book-PL-IND
        'a book'   'three books'   'some books'

Klaus von Heusinger & Roya Sadeghpoor

combined, yielding three different indefinite configurations: *ye(k) NP*, *NP-i*, and *ye(k) NP-i*. For Standard Written Persian, the suffixed *-i* has indefinite readings, including readings that undergo negation and other operators. The use of *ye(k)* is thought to express the typical "cardinal" reading of indefinites. There is no clear delimitation of the function of the combined form *ye(k) NP-i*.

Windfuhr (1979) considers *NP-i* to have three functions: (i) as *-i* of 'unit', the construction has similar functions as *a(an)* in English; (ii) as *-i* of indefiniteness, the construction is very similar to what Jasbi (2016) describes as 'antidefinite', similar to 'any' or 'some'; (iii) as demonstrative *-i*, the construction appears with relative clauses.[5] Toosarvandani and Nasser (2017) report that some traditional (Lambton 1953) as well as contemporary linguists (Ghomeshi 2003) assume that the indefinite determiner *yek+NP* and the suffixed *NP-i* can be equivalent in positive, assertive contexts, see example (16) (mainly in non-contemporary or more literary usages); however, Toosarvandani and Nasser (2017) provide examples that show a difference in distribution and meaning between the two constructions, mainly in negative, non-assertive contexts, see examples (17) and (18). In the following, the two indefinites' similarities and differences are discussed.

Since *-i* is a suffix, it can occur with quantifiers. In fact, when universal quantifiers such as *har* ('every') and *hich* ('no') are present, the suffixed *-i* usually accompanies the NP. Lyons (1999: 90) states that the "suffix *-i* semantically marks the noun phrase as non-specific or arbitrary in reference and is approximately equivalent to *any* in nonassertive contexts and *some...or other* in positive declarative contexts". Ghomeshi (2003: 64-65) argues that the two forms partly overlap, but that the suffixed *-i* has a wider range of application. She does not discuss the combined form, however. Paul (2008: 325) argues that *-i* has the function of "picking out and individuating entities". He argues that this function should be kept separate from specificity and referentiality. Hincha (1961: 169-170) assumes that *ye(k)* expresses an individualized entity, while *-i* signals an arbitrarily chosen element of a class. Modarresi (2014: 16-19) focuses on the differences between bare nouns in an object position, and *ye(k) NP* and *NP-i* objects. The latter both introduce discourse referents and show scopal effects, while the bare noun does not. We cannot do justice to the whole discussion on indefinites in written Persian, but we try to summarize the main, and hopefully uncontroversial, observations in Table 2.

---

[5]There is an ongoing discussion as to whether the use with relative clauses is a use of the suffixed article or a different morpheme (see discussion in Ghomeshi 2003: 65).

Table 2: Definite and indefinite constructions in Standard Written Persian

Function	Form	Positive context	Negative context
definite	*NP*	uniqueness and familiarity	wide scope
indefinite – expressing 'cardinality' or 'existential entailment'	*ye(k) NP*	cardinal reading, existential 'a/one N'	variation between wide and narrow scope
indefinite, existential	*NP-i*	existential/'one of a class' 'some or other'	NPI-narrow scope-'any'
indefinite – expressing 'cardinality' or 'existential entailment' and speaker's indifference or ignorance	*ye(k) NP-i*	existential/arbitrary in reference/speaker's indifference similar to German: '(irgend) jemand'	wide scope

Semantically, *yek NP-i* can express existence and signals that the referent is arbitrarily chosen (Lyons 1999). Pragmatically, it can show a speaker's indifference or ignorance, or a free-choice implication (Jasbi 2016). In written form, the three indefinites behave similarly in positive declarative contexts, as shown in (16a-c).

(16)   Context: There were three books. Ali bought one of them.

    a. Ali ketāb-i xarid.
       Ali book-i bought.3sɢ
       'He bought a/some book.'

    b. Ali yek ketāb xarid.
       Ali a   book bought.3sɢ
       'He bought a book.'

    c. Ali yek ketāb-i xarid.
       Ali a   book-i bought.3sɢ
       'Ali bought a/some book.'

Considering negation, *ye(k) NP-i* takes a wide scope over negatives and questions, while *NP-i* takes a narrow scope in the same context, and *ye(k) NP* can take variable scope (Toosarvandani & Nasser 2017: 8-9; Modarresi 2014: 26-30). The acceptability of a wide scope under negation with different indefinites is illustrated in (17) and (18). Context (17) forces a narrow-scope reading for the indefinites, which is available for *NP-i* in (17a) and *ye(k) NP* in (17b), but not for

Klaus von Heusinger & Roya Sadeghpoor

ye(k) NP-i in (17c). The context in (18) strongly suggests a wide-scope reading, which is not available for *NP-i* in (18a), but possible for *ye(k) NP* in (18b), and for *ye(k) NP-i* in (18c). (Note that the wide-scope reading goes hand in hand with the object marker *-rā*.)

(17)   Context: There were three possible books I could buy. I didn't buy any of them.

    a.   Man **ketāb-i** naxaridam.
        I   **book-i** not.bought.1SG
        'I didn't buy any books.'   ¬ > ∃

    b.   Man hattā **ye ketāb** ham naxaridam.
        I   even **a book** also not.bought.1SG
        'I didn't buy any book.'   ¬ > ∃

    c.   #Man **ye ketāb-i** ro naxaridam.
        I   **a book-i** rā not.bought.1SG
        Intended: 'I didn't buy any book.'   *¬ > ∃

(18)   Context: There were three possible books that I could buy. I bought two of them but not the third.

    a.   #Man **ketāb-i** naxaridam.
        I   **book-i** not.bought.1SG
        'There is a book I didn't buy.'   *∃ > ¬

    b.   Man **ye ketāb** ro naxaridam.
        I   **a book** rā not.bought.1SG
        'There is a book I didn't buy.'   ∃ > ¬

    c.   Man **ye ketāb-i** ro naxaridam.
        I   a **book-i** rā not.bought.1SG
        Intended: 'There is a book I didn't buy.' [6]   ∃ > ¬

    As shown in (18a), *NP-i* takes wide scope neither under negation nor under questions (similar to NPIs). However, in positive contexts (written form), it behaves similarly to simple indefinites and can have an existential or numerical implication.

---

[6](18c) is felicitous in the written variety with DOM 'ra' whereas it is not felicitous in Modern Colloquial Persian.

3.3 Indefiniteness marking in Modern Colloquial Persian

One of the main distinctions between the system of indefinite forms in the written vs. spoken register is the semantic role of suffixed *-i*. In the written register, *-i* is a common way of marking an indefinite NP, whereas in colloquial Persian, *yek NP* is common and *-i* is very restricted as it is used as an NPI. Jasbi (2016: 246) categorizes the indefinite markers in his native Tehrani colloquial Persian into three main categories: simple, complex, and antidefinite. He illustrates their difference in the following table:

Table 3: Definite and indefinite constructions in Modern Colloquial Persian (Jasbi 2016: 246)

	Type	Form	Example	Translation
Definites	Bare	*NP*	*māshin*	the car
Indefinites	Simple	*ye NP*	*ye māshin*	a car
	Antidefinite	*NP-i*	*māshin-i*	∼ a/any car
	Complex	*ye NP-i*	*ye-māshin-i*	∼ some car or other

Jasbi calls *ye(k) NP* a simple indefinite because it behaves similarly to *a(n)* in English and carries an existential inference ($|[\![NP]\!]| \geq 1$). On the other hand, *NP-i* entails an antidefinite interpretation, meaning that it rejects any set that can have a unique inference ($|[\![NP]\!]| \neq 1$) and can have a non-existential implication ($|[\![NP]\!]| = 0$). Therefore, the respective set either is empty or contains more than one element. Now, the complex indefinite *ye(k) NP-i* has an anti-singleton implication ($|[\![NP]\!]| > 1$), which is compositionally derived from the existential inference and the anti-uniqueness condition. The summary of the semantic differences proposed by Jasbi (2016: 251) is provided in Table 4.

Table 4: Cardinality implications for definites and indefinites in Modern Colloquial Persian (Jasbi 2016: 251)

	Type	Form	Cardinality
Definite	Bare	*NP*	$|[\![NP]\!]| = 1$
Indefinite	Simple	*ye NP*	$|[\![NP]\!]| \geq 1$
	Antidefinite	*NP-i*	$|[\![NP]\!]| \neq 1$
	Complex	*ye NP-i*	$|[\![NP]\!]| > 1$

To summarize, the function of the different indefinite markers in Standard Written Persian is controversial, and their function in Standard Colloquial Persian requires more investigation. Based on Jasbi's (2016) semantic characterization (see Table 4) and the examples discussed above as well as in the subsequent sections, we assume that the form *ye(k) NP* corresponds to the unmarked indefinite, the form *NP-i* only appears with negation, in conditionals, and in questions, and the combined form *ye(k) NP-i* expresses a speaker's ignorance or indifference.

4 The specificity marker *-e* in Modern Colloquial Persian

Modern Colloquial Persian has the suffix *-e*, which can optionally combine with bare, i.e., definite, noun phrases, demonstrative noun phrases, and indefinite noun phrases. With bare noun phrases, *-e* is assumed to express a demonstrative or definite function (Windfuhr 1979: 40; Lazard 1957: 163; Ghomeshi 2003: 67; Toosarvandani & Nasser 2017; Jasbi 2020a). Hincha (1961: 173-177) summarizes the distributional properties of *-e*: it is always optional — there are no conditions that makes its use obligatory. If used, it is always accented and attached directly to the stem. It stands in opposite distribution to the plural suffix *-hā*, i.e., either *-hā* or *-e* can be used, but not both, which leads Hincha (1961: 175) to assume that both suffixes share some features and express some contradictory features, such as number. Ghomeshi (2003: 68) adds that *-e* "cannot attach to anything already of category D", such as proper names, pronouns, and noun phrases containing possessors. It cannot combine with the suffixed marker *-i*, but as we will discuss below, it can combine with the complex *ye(k) NP-i*. With indefinite noun phrases, the suffix signals specificity. In the following, we first provide an overview of specific definite contexts that license the use of the marker, and then provide data on the possible combination of the marker with indefinite constructions.

4.1 Specificity marker *-e* with definites

Modern Colloquial Persian can express (certain kinds of) definiteness by means of the marker *-e*, which is absent in Standard Written Persian (Windfuhr 1979: 50; Ghomeshi 2003). The function of *-e* is described as demonstrative, definite, determinative, or referential. Hincha (1961: 176) assumes that *-e* signals that the NP refers to one particular or individualized entity ("Einzelgegenstand"). There is no comprehensive study of this marker.

There is an interesting distribution of *-e* with the unmarked bare noun. Nikravan (2014) argues that there is a functional difference between unmarked

noun phrases, on the one hand, and noun phrases marked with *-e* on the other. The former express weak definiteness and the latter strong definiteness, as is found in other languages with two definite articles (see Schwarz 2013). Strong forms are used in anaphoric and situational contexts, while weak forms appear in encyclopedic, unique, and generic contexts. This contrast is illustrated in (19).

(19) Emruz yek pesar va yek doxtar ro didam. Pesar??(-e) ro mišnāxtam.
today a boy and a girl rā saw.1SG boy??(-e) rā knew.1SG
'Today I saw a boy and a girl. I knew the boy.'

In (19), *pesar* 'boy' is anaphoric and much more acceptable with the marker *-e* than without it. Consequently, it is argued that in contexts where an explicit antecedent is present, the strong definite is used. Other scholars propose that *-e* marks familiarity of the associated referent (Hedberg et al. 2009). The results of a questionnaire presented in Nikravan (2014) seem to indicate that there is a marginal effect of *-e* towards a familiarity reading. However, it is unclear from her presentation whether the effect is statistically reliable or not. The results also show that the use of *-e* is optional, as in (19).

The use of *-e* with different types of definite noun phrases (see (14) above) provides further evidence that (i) the use of *-e* is optional and (ii) *-e* can only be used with referential definites, i. e., anaphorically used definites, as in (20a), and definites in bridging contexts, as in (20b). The use of *-e* is ungrammatical for unique definites, as in (20c), and generic uses, as in (20d).

(20) a. Anne yek xune xaride.   **Xune(-he)** labe marze kešvare.
Anne a house bought.3SG **house(-he)** on edge.of country.be.3SG
'Anne bought a house. The house is at the border of the country.'

b. Anne rafte   bud   ye marāseme arusi.
Anne went.3SG AUX.3SG a ceremony marriage

Arus(-e) xeyli xošgel bud.
bride(-e) very beautiful was.3SG
'Anne went to a marriage ceremony. The bride was very beautiful.'

c. **Māh(*-e)** xeyli rošan miदेraxše.
moon(*-e) very bright PROG.shine.3SG
'The moon shines very brightly.'

d. **Dianāsor(*-e)** 60 milion sāle qabl monqarez šode.
dinosaur(*-e) 60 million year ago extinct became.3SG
'Dinosaurs became extinct 60 million years ago.'

The referential function of *-e* can also be shown in the contrast between a referential and an attributive reading of a definite NP (Donnellan 1966; Keenan & Ebert 1973). Sentence (21) strongly suggests an attributive or non-referential reading of the noun *barandeye* 'the winner, whoever the winner will be'. In this reading, the use of *-e* is ungrammatical, which confirms the assumption that *-e* signals referentiality, in the sense that the hearer, as well as the speaker, can uniquely identify the referent of the noun phrase.

(21) **Barandeye(*-e)** in   mosābeqe   yek safar be ālmān   migirad.
       winner(*-e).of this competition a   trip to Germany get.3sG
       'The winner of this competition (whoever he/she is) will get a trip to Germany.'

Therefore, we can conclude that the function of *-e* is to mark referentially strong definites, i. e., definites that refer to a discourse referent that was explicitly or implicitly introduced into the linguistic context.

4.2 The suffix *-e* with indefinites

The specificity marker *-e* can combine with two of the three indefinite configurations, as in the examples (6)-(8) above, repeated here as (22)-(24).

(22) **Ye(k) pesar-e** injā kār   mikone.
       a   boy-e   here work do.3sG
       'A (specific) boy works here.'

(23) Diruz   **māšin-i** jole   dare   xune didi?
       yesterday **car-i**   front.of door.of home saw.2sG

       – *Na, man hič   **māšin-e-i** nadidam.
         no I   any **car-e-i**   not.saw.1sG
       'Did you see any cars in front of the house door yesterday?'
       Intended: 'No, I didn't see any specific car.'

(24) Emruz **ye māšin-e-i** az   pošt   behem zad.
       Today **a car-e-i**   from behind to.me smashed.3sG
       'Today some/other (specific) car smashed me from behind.'

The form *NP-i* cannot combine with *-e*. We speculate that this is due to a conflict of the referential meaning of *-e* and the NPI-meaning of *NP-i* in Modern Colloquial Persian.[7]

---

[7]The occurrence of *-e* with *NP-i* is not possible with restrictive relative clauses (Ghomeshi 2003).

However, both forms with the indefinite article *ye(k)* can combine with *-e*, yielding *ye(k) NP-e* and *ye(k) NP-e-i*, respectively. With both indefinite constructions, the marker *-e* signals referential and wide-scope readings of the indefinites. The regular indefinites *ye doxtar* in (25a) and *ye doxtar-i* in (25c) allow for (i) a wide-scope and (ii) a narrow-scope reading with respect to the universal quantifier. However, the forms *ye doxtar-e* in (25b) and *ye doxtar-e-i* in (25d) only allow for a wide-scope, referential, or specific reading. We find the same contrast for indefinites in sentences with verbs of propositional attitudes, as in (26). The *-e* marked indefinites can only take a wide scope with respect to the intensional operator *mixad* 'to want'.

(25)   a. Hame pesar-hā bā   ye doxtar raqsidan.
           all   boy-PL with a girl   danced.3PL
           (i) 'There is a girl such that every boy danced with her.'
           (ii) 'For every boy, there is a different girl, such that, that boy dances with her.'

      b. Hame pesar-hā bā   ye doxtar-e raqsidan.
           all   boy-PL with a girl-e danced.3PL
           (i) 'There is a girl such that every boy danced with her.'

      c. Hame pesar-hā bā   ye doxtar-i raqsidan.
           all   boy-PL with a girl-i danced.3PL
           (i) 'There is a girl such that every boy danced with her.'
           (ii) 'For every boy, there is a different girl, such that, that boy dances with her.'

      d. Hame pesar-hā bā   ye doxtar-e-i raqsidan.
           all   boy-PL with a girl-e-i danced.3PL
           (i) 'There is a girl such that every boy danced with her.'

(26)   a. Ali mixād   bā ye doxtar dust   še.
           Ali want.3SG with a   girl   friend become.3SG
           (i) 'Ali wants to make friends with a specific girl.'
           (ii) 'Ali wants to make friends with a girl/whoever she may be.'

      b. Ali mixād   bā ye doxtar-e dust   še.
           Ali want.3SG with a   girl   friend become.3SG
           (i) 'Ali wants to make friends with a specific girl.'

    c. Ali mixād   bā   ye doxtar-i dust   še.
        Ali want.3sg with a   girl-i   friend become.3sg
        (i) 'Ali wants to make friends with a specific girl.'
        (ii) 'Ali wants to make friends with a girl/whoever she may be.'

    d. Ali mixād   bā   ye doxtar-e-i dust   še.
        Ali want.3sg with a   girl-e-i   friend become.3sg
        (i) 'Ali wants to make friends with a specific girl.'

We take the distribution of *-e* discussed here as a good evidence that the marker encodes a specific or referential reading of the indefinite.[8]

4.3 Specificity marker and referential anchoring

Epistemic specific indefinites express the "referential intention" of the speaker. That is, the speaker signals with these expressions that he or she has already decided on the referent of the indefinite. Non-specific indefinites, on the other hand, assert the existence of an individual that falls under the descriptive content of the indefinite, but not a particular individual. The concept of epistemic specificity as speaker-oriented (or speaker-anchored) seems too narrow, however, as we also find (epistemic) specific indefinites where the speaker cannot identify the referent, but can recognize some other salient discourse participant. Therefore, von Heusinger (2002; 2019) proposes the concept of "referential anchoring", modeling the dependency of the referent of the indefinite from some other salient discourse referent or participant (typically the speaker, the subject of the sentence, etc.). The discussion of the contrast between the specificity adjectives *ein gewisser* and *ein bestimmter* in section 2.2 was explained along these lines: *ein gewisser* is speaker-oriented, while *ein bestimmter* is not obligatorily speaker-oriented, i. e., it can also be anchored to another salient agent in the discourse.

---

[8]Here we leave open what the exact semantics of the marker *-e* is. Hincha (1961: 176) describes it as "punctualization"; Jasbi (2020b) assumes that the marker *-e* creates a singleton set, thereby simulating wide-scope behavior. However, this approach would not explain why it can be used with certain definites and why it can be combined with the complex form *ye(k) NP-e-i*, as it would include the combination of a singleton and an anti-uniqueness condition. An alternative approach is to assume that the marker is interpreted as an indexed choice function (Egli & von Heusinger 1995; Winter 1997) that selects one element out of a set. This would explain the use with certain definites, and also the complementary distribution with the plural suffix *-hā*. Such an account could provide an explanation for the definiteness effect on bare nouns. (The value for the index of the choice function is provided by the local situation or the local discourse, but not by encyclopedic knowledge.) In the form *ye(k) NP-e*, the index is locally bound by the speaker, and for the form *ye(k) NP-e-i*, the index can also be bound by other salient discourse referents.

The two indefinite forms *ye(k) NP-e* and *ye(k) NP-e-i* are interpreted as specific or referential indefinites. We suggest that the difference between the two forms lies in the specificity orientation in epistemic contexts. It seems that the form *ye(k) NP-e-i* is less acceptable in general; however, we still find examples such as the following on Twitter:[9]

(27) Tanhāi vasate pārke Mellat nešastam, ye xānum-e-i dāre
      alone   middle.of park.of Mellat sitting.1SG a woman-e-i AUX.3SG
      kenāram Qurān mixune.
      next.me Quran reading.3SG
      'I am sitting alone in the middle of Mellat Park and some woman is reading the Quran next to me.'

(28) Ye doxtar-e-i tu bašgāham hast, ajab heykali dāre!
      a girl-e-i at gym.my be.3SG what body-shape have.3SG
      'There is some girl at the gym I go to, that has a perfect body-shape!'

(29) Zošk ke berim, ye sag-e-i ham hast unjā, tāze āšnā šode
      Zoshk that go.1PL a dog-e-i also be.3SG there recently familiar got.3SG
      bāhām.
      with.me
      'If we go to Zoshk, there is some dog there that made friends with me last time.'

(30) Ye dars dāštam be esme "Riāzi Pišrafte". Unjā ye pesar-e-i
      a course had.1SG with name "Math advanced" there a boy-e-i
      bud   be esme Vahid ya Hamid.
      was.3SG with name.of Vahid or Hamid
      'I had a course called "Advanced Mathematics". There was some boy named Vahid or Hamid.'

We propose that the basic function of the suffixed indefinite article *-i* in Modern Colloquial Persian is to signal speaker ignorance or indifference. Combining speaker ignorance with the epistemic specificity or referentiality might

---

[9]The first anonymous reviewer pointed out that all the Twitter examples (27)-(30) are speaker-oriented and would therefore contradict our hypothesis that the form *ye(k) NP-e-i* is non-speaker-oriented. We think that it is difficult to judge this without more context. Moreover, we believe that, in most of the examples, the speaker signals that he or she is not able or willing to reveal the identity of the indefinite NP. The main point of the Twitter examples is to show that these forms are in current use, which contradicts some assumptions made in the literature.

Klaus von Heusinger & Roya Sadeghpoor

result in a semantic-pragmatic condition which we have termed non-speaker-oriented specificity (see discussion in section 2.2 above). Therefore, we hypothesize that the difference between these two forms is the orientation or anchoring of the specificity relation. For *ye(k) NP-e*, we assume that the indefinite is referentially anchored to the speaker, i.e., the indefinite is speaker-oriented specific. The form *ye(k) NP-e-i*, in contrast, is referentially anchored to a discourse referent other than the speaker, i.e., it is non-speaker-oriented. We summarize this hypothesis in Table 5.[10]

Table 5: Specificity marker *-e* with different indefinite markers in Modern Colloquial Persian

Indefinite form	Cardinality (Jasbi 2016)	Pragmatic difference	Combination with specificity marker *-e*	Assumed function
ye(k)+NP	$|[\![NP]\!]| > 1$	normal indefinite marker like *a(n)*	acceptable	speaker-oriented specificity
NP-i	$|[\![NP]\!]| \neq 1$	with negations, conditionals, and questions	ungrammatical	–
ye(k)+NP-i	$|[\![NP]\!]| > 1$	speaker's indifference or ignorance is more likely	less acceptable	non-speaker -oriented specificity

Our hypothesis makes clear predictions about the acceptability of sentences containing these forms in contexts that express a speaker orientation vs. a non-speaker orientation of the indefinite. We assume that the indefinite *ye ostād-e* expresses a speaker orientation, which predicts that the continuation (31i) is coherent, while the continuation (31ii) is incoherent. For the indefinite *yek ostād-e-i* in sentence (32), we assume a non-speaker-orientation, which predicts that continuation (32i) is not felicitous, while continuation (32ii) is.

---

[10]Our second reviewer asks whether we assume a compositional semantics, which would provide an independent function for each marker, or whether we assume just one function for the whole construction. For a compositional approach, see Jasbi (2016) for the indefinite forms, and footnote 8 in this chapter, on the choice function approach to the specificity marker *-e*. However, we have not yet developed full semantics for these configurations.

(31) Sārā **ye ostād-e**   **ro** xeyli dust dāre.
      Sara **a   professor-e rā** very like have
      'Sara likes a specific professor very much.'

  (i) Man midunam kudum ostād.
       I   know.1sg which professor
       'I know who he is.'

  (ii) #Vali nemidunam   kudum ostād.
        but not.know.1sg which professor
       'But I don't know which professor.'

(32) Sārā **ye ostād-e-i**   **ro** xeyli dust dāre.
      Sara **a   professor-e-i rā** very like have
      'Sara likes some specific professor very much.'

  (i) #Man midunam kudum ostād.
        I   know.1sg which professor
        'I know who he is.'

  (ii) Vali nemidunam   kudum ostād.
       but not.know.1sg which professor
       'But I don't know which professor.'

We can summarize this prediction in Table 6 with the expected acceptability of the continuation.[11]

---

[11]The second reviewer also suggested that we test the examples (31)-(32) without the specificity marker *-e*, as in (31′) and (32′). The reviewer reported that his or her informants would accept the continuations (i) and (ii) for both sentences, but that the informants expressed a preference for (31′ii) and (32′i), which would be the opposite of the expectation expressed for (31)-(32). We agree that both continuations are good for both sentences, but we do not share their preferences. We do not have any predictions with respect to (31′) and (32′). Note that both (31)/(32), and (31′)/(32′), have the direct object marker *-rā*, which is assumed to express specificity by itself. We cannot go into details about the difference between the function of *-e* and *-rā* here; however, our test items had examples with and without *-rā*.

(31′) Sārā **ye ostād**   **ro** xeyli dust dāre.   (i) Man midunam kudum ostād.
       Sara **a   professor rā** very like have   I   know.1sg which professor
       'Sara likes a specific professor very much.'   'I know who he is.'

(32′) Sārā **ye ostād-i**   **ro** xeyli dust dāre.   (ii) Vali nemidunam   kudum ostād.
       Sara **a   professor-i rā** very like have   but not.know.1sg which professor
       'Sara likes some specific professor very much.'   'But I don't know which professor.'

Table 6: Prediction of type of epistemic specificity of *-e* marked indefinites

Indefinite form	Epistemic type	Example	Context	Predicted acceptability
ye(k) NP-e	speaker-specific	(31)	(i) speaker knowledge	very good
ye(k) NP-e	speaker-specific	(31)	(ii) speaker ignorance	#
ye(k) NP-e-i	non-speaker-specific	(32)	(i) speaker knowledge	#
ye(k) NP-e-i	non-speaker-specific	(32)	(ii) speaker ignorance	good

5 Empirical evidence for speaker orientation of specific noun phrases

In this section, we present two pilot acceptability studies that tested the predictions outlined above. In the first pilot, we used eight sentences, which we continued with either (i), a context that was only coherent with a speaker-specific reading or (ii) a context that was only coherent with a non-speaker-oriented specific reading. The results show that simple indefinites with *ye(k) NP-e*, regardless of their specificity orientation, are more acceptable than complex indefinites, but there were no clear effects of specificity orientation. We assume that our results might reflect a mix-up between different degrees of animacy in the included indefinites. Therefore, we conducted a second pilot study with only human indefinites and a different design; as well as simple sentences and their speaker-oriented vs. non-speaker-oriented continuations, we also presented sentences that clearly signaled speaker ignorance in order to test whether informants can distinguish between different specificity orientations. The results of the second study not only confirm that speakers are capable of making this distinction, but also provide some support for our claim that the simple indefinite *ye(k) NP-e* is speaker-oriented, and the complex indefinite *ye(k) NP-e-i* is non-speaker-oriented.

5.1 Experiment 1

Our hypothesis H1 is that in Modern Colloquial Persian *ye(k) NP-e* always functions as speaker-specific ('gewiss NP'), while *ye(k) NP-e-i* can only function as non-speaker-oriented. In order to test this hypothesis, we conducted a pilot questionnaire with speakers of Modern Colloquial Persian. We used a simple sentence, as seen in (33), with simple indefinites with the marker *-e* (*yek doktor-e*), as well as complex indefinites with the marker *-e* (*yek doktor-e-i*). The first sentence with the

critical item (*yek doktor-e* or *yek doktor-e-i*) is continued with either (i) an assertion that the speaker had knowledge of the referent, or (ii) a statement signaling the ignorance of the speaker. That is, continuation (i) strongly forces a speaker-specific reading and continuation, while (ii) forces a non-speaker-specific reading. Note that we did not test indefinites without the marker *-e*, as we assume that there is ambiguity between a specific and non-specific interpretation.[12]

(33) a. Mona bā   **yek doktor-e** ezdevaj   karde.
          Mona with **a**   **doctor-e** marriage did.3SG
          'Mona married a doctor.'

      b. Mona bā   **yek doktor-e-i** ezdevaj   karde.
          Mona with **a**   **doctor-e-i** marriage did.3SG
          'Mona married a doctor.'

       (i) Man midunam kudum doktor.
            I   know.1SG which doctor
            'I know which doctor he is.'

      (ii) Vali man nemidunam   kudum doktor.
            but I   not.know.1SG which doctor
            'But I do not know which doctor he is.'

5.1.1 Participants and experimental technique

Twenty male and female participants participated in the study. Their native language was Persian and they had lived all or most of their lives in Iran. Their ages varied between 25 and 67. In terms of educational level, six participants had high school diplomas, ten had bachelor's degrees, and four had master's degrees. Participants read Persian written texts for at least one hour a day, and they spoke/heard Persian all or most of the day.

The study followed a 2x2 design with two different indefinite forms: (a) *ye(k) NP-e* and (b) *ye(k) NP-e-i* and two continuations: (i) "I do know who/which" for the speaker-oriented epistemic specificity and (ii) "I do not know who/which" for the non-speaker-oriented epistemic specificity. The assumption was that all forms are epistemically specific, as in Table 6 above. We had eight different sentences and

---

[12]In half of the examples the critical indefinite was the direct object, as in (31), and a different argument in the other half, as in (33). We found that this alternation had no significant effect, even though we added the differential case marker *-rā* in the direct object instances. It is unclear what additional function this marker performs (see the discussion in the last footnote). We also balanced for animacy, see the discussion below and Figure 2.

Klaus von Heusinger & Roya Sadeghpoor

created four lists using a Latin square design, so that each participant heard one sentence and two conditions each. Probable factors which might intervene with the evaluation, such as animacy, position of NP in the sentence (direct object/indirect object), and direct/indirect speech, were equally present in all items.

As we were testing Modern Colloquial Persian, i. e., spoken Persian, we read out the sentences to our participants at least once and asked them to evaluate the sentence on a scale from 1 for "completely acceptable" to 7 for "completely unacceptable" on the answer sheet, where they were also able to read the test sentence themselves.

5.1.2 Results

We observed that participants complained (even verbally) about the appearance of *-e* in *ye(k) NP-e-i* in both speaker-specific and non-speaker-specific readings. This is also reflected in the acceptability scores. We summarize the pilot questionnaire with 20 participants in Table 7, together with the expected acceptability.

Table 7: Effect of *-e* as specificity marker of indefinites on the kind of epistemicity (1 = very good; −7 = very bad)

| Indefinite form	Epistemic type	Mean value	Acceptability	Predicted acceptability
ye(k) NP-e	speaker-specific	2.725	good	very good
ye(k) NP-e	non-speaker-specific	3.025	good	#
ye(k) NP-e-i	speaker-specific	4.675	weak	#
ye(k) NP-e-i	non-speaker-specific	4.425	weak	good
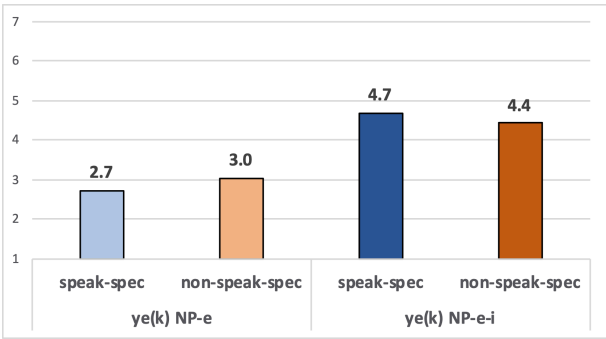
Figure 1: Acceptability (1 = very good −7 = very bad) of simple and complex indefinites in non-speaker and speaker-oriented specificity contexts

138

Overall, we see that the form *ye(k) NP-e* was more acceptable than the form *ye(k) NP-e-i*, which confirms the intuition reported above. However, we also see that *ye(k) NP-e* performed well in both conditions (speaker- and non-speaker-specific), which went against our hypothesis. The judgment for the non-speaker-specificity condition is marginally weaker. The form *ye(k) NP-e-i* was clearly weaker; however, there is only a marginal difference between speaker-oriented specificity (slightly weaker) and non-speaker-oriented specificity. Interestingly, when distinguishing between human and non-human indefinites, as in Figure 2, we see that the non-human indefinites were less acceptable than the human indefinites. Furthermore, when looking at the human indefinites we can see that the simple indefinites (*ye(k) NP-e*) were rated as slightly better in the speaker-specificity condition than in the non-speaker conditions (1.85 vs. 2.35). Complex indefinites (*ye(k) NP-e-i*), on the other hand, were slightly better in the non-speaker-specificity condition than in the speaker-oriented specificity condition (3.9 vs. 4.3).
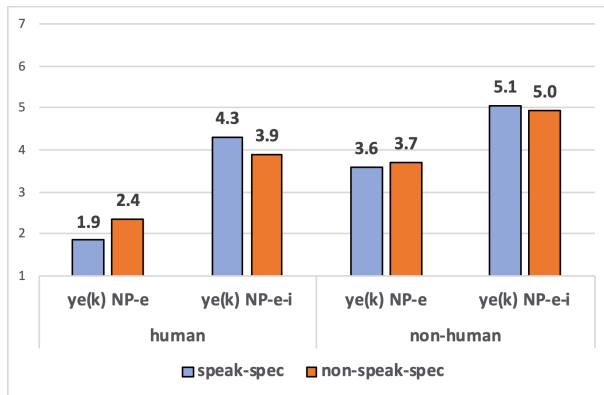
Figure 2: Acceptability (1 = very good; −7 = very bad) of simple and complex indefinites for human and non-human noun phrases in non-speaker and speaker-oriented specificity contexts

5.1.3 Discussion

Our first pilot study shows that complex indefinites with the marker *-e* are less acceptable than simple indefinites with the marker *-e*. Animacy is also an important factor: our study demonstrates that human indefinites were more acceptable than non-human indefinites. However, the predicted contrast between simple and complex indefinites in speaker- vs. non-speaker-oriented specificity contexts was not shown to be significant. We surmise that this contrast might be more pronounced with human indefinites, which led us to design a second pilot experiment.

Klaus von Heusinger & Roya Sadeghpoor

5.2 Experiment 2

In order to test the hypothesis that the two specific indefinites in Modern Collo-
quial Persian differ with respect to the referential anchoring of the indefinite, i. e.,
in the specificity orientation, in the second study, we focused on human indefi-
nites. Additionally, we included some examples that provided contexts that sig-
naled speaker ignorance in the first sentence. These examples were used to test
whether participants were sensitive to the speaker- vs. non-speaker-orientation.

5.2.1 Design

Experiment 2 was conducted to test for a feature which is only present in spo-
ken colloquial Persian, namely the *-e* marker with indefinite NPs. It followed
the same 2x2 design with four lists as the first pilot study. There were 24 items
consisting of 12 test and 12 filler items in each list. The experimental stimuli con-
sisted of two sentences for each item. Since the feature under investigation was
simple vs. complex indefinites with the marker *-e*, the first sentence contained an
indefinite noun phrase either with *yek NP-e* or *yek NP-e-i*. The second sentence
forced either a speaker-specific reading of the indefinite in the first sentence, or
a non-speaker-specific reading.

   In the speaker-specific continuation, we asserted the knowledge of the
speaker about the identity of the referent of the indefinite. In the non-speaker-
specific continuation, we asserted the ignorance of the speaker about the identity
of the referent, thus forcing a non-speaker-specific reading.

(34)   Critical items for Experiment 2

   a. Simple indefinite (***yek NP-e***) + speaker-specific continuation

   Sara emruz az   **ye vakil-e** vaqte   mošāvere gerefte.
   Sara today from **a lawyer-e** appointment consulting took.3sɢ

   Man ham ba   vakil-e čandinbar   kar kardam.
   I   also with lawyer-e several.time work did.1sɢ

   Kareš   xeyli xube.
   work.his very good.be.3sɢ

   'Sara had a consultation appointment with a lawyer today. I have also
   consulted with the lawyer. His work is very good.'

b. Simple indefinite (**yek NP-e**) + non-speaker-specific continuation

Sara emruz az   **ye vakil-e**   vaqte   mošāvere gerefte.
Sara today from **a   lawyer-e** appointment consulting took.3sg

Migan vakil-e   maroofe   vali man čiz-i   azaš
say.3pl lawyer-e known.be.3sg but I   thing-INDEF from.him
nemidunam.
not.know.1sg

'Sara had a consultation appointment with a lawyer today. They say
that this lawyer is well known, but I do not know anything about him.'

c. Complex indefinite (**yek NP-e-i**) + speaker-specific continuation

Sara emruz az   **ye vakil-e-i**   vaqte   mošāvere gerefte.
Sara today from **a   lawyer-e-i** appointment consulting took.3sg

Man ham ba   vakil-e   čandinbār   kār   kardam.
I   also with lawyer-e several.time work did.1sg

Kareš   xeyli xube.
work.his very good.be.3sg

'Sara had a consultation appointment with a lawyer today. I have also
consulted with the lawyer, several times. His work is very good.'

d. Complex indefinite (**yek NP-e-i**) + non-speaker-specific continuation

Sara emruz az   **ye vakil-e-i**   vaqte   mošāvere gerefte.
Sara today from **a   lawyer-e-i** appointment consulting took.3sg

Migan vakil-e   maroofe   vali man čiz-i   azaš
Say.3pl lawyer-e known.be.3sg but I   thing-INDEF from.him
nemidunam.
not.know.1sg

'Sara had a consultation appointment with a lawyer today. They say
that this lawyer is well known, but I do not know anything about him.'

The test items also differed in their constructions: eight items had a third
person subject (proper name), as in (34), and four other items were constructions
that showed a greater distance from the speaker, namely two items of the type
"They say…", as in (35), and two items of the form "I heard…", as in (36). Note that
we provide only the a-condition with *ye(k) NP-e* and the specific continuation,
as in (34a).

Klaus von Heusinger & Roya Sadeghpoor

(35) "They say..." construction:

Migan **ye moalem-e** tu madrese Tizhushān hast   ke   bečeha
say.3PL **a   teacher-e** in school   Tizhushan be.3SG that student.PL
azaš   xeyli mitarsan.   Man ham bāhāsh 4ta dars   daštam va
from.her very frighten.3PL I   also with.him 4CL course had.3SG and
oftadam.
failed

'They say there is a teacher in Tizhushan school that every student is
afraid of. I also have had four courses with him and failed them all.'

(36) "I heard..." construction:

Šenidam **ye pesar-e** hast   tu in   mahale
heard.1SG **a   boy-e**   be.3SG in this neighborhood

ke   vase doxtara mozāhemat ijad   mikone.
who for   girl.PL harassment make do.3SG

Man mišnasameš az   vaghti bače bud.
I   know.1SG   him.from when   child was.3SG

'I have heard that there is a boy in this neighborhood who harasses girls.
I have known him since he was a child.'

5.2.2 Results of Experiment 2

There was strong agreement in relation to the filler/control items, with marginal
differences between participants (< 0.8 points). The results of the test items can
be summarized as follows. Firstly, in contrast to Experiment 1, Figure 3 does not
show a clear preference for simple indefinites. Rather, both types were rated
very similarly. Secondly, we clearly see that the contexts which signaled speaker
ignorance ("They say... ", "I heard...") preferred non-speaker-oriented specificity
continuations. It shows that participants were aware of this contrast.
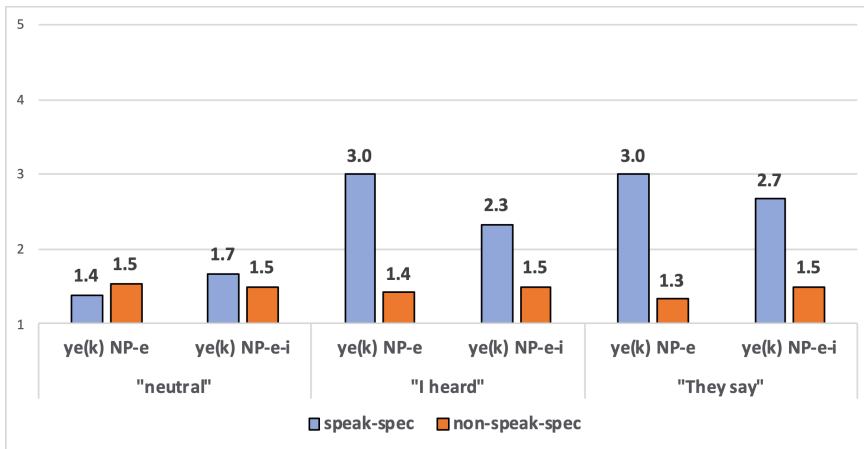
Figure 3: Acceptability (1 = very good; –5 = very bad) of simple and complex indefinites for human noun phrases in non-speaker and speaker-oriented specificity contexts, across types of constructions

A more detailed inspection of the neutral contexts in Figure 3 reveals a slight preference for the simple indefinite *ye(k) NP-e* in speaker-oriented specificity contexts (1.39) vs. non-speaker-oriented specificity contexts (1.54), while the complex indefinite *ye(k) NP-e-i* was rated slightly better in non-speaker-oriented specificity contexts (1.5) vs. speaker-oriented specificity contexts (1.67).

In summary, the direct comparison in neutral contexts between the simple and the complex indefinite with the marker *-e* does not provide significant contrasts. It only suggests a preference of the simple indefinite for speaker-oriented specificity, while the complex indefinite prefers non-speaker oriented specificity. However, constructions with "They say..." or "I heard...", which clearly encode non-speaker oriented specificity, show a preference for the complex indefinite. This supports our hypothesis for the difference between the two specific indefinites.

6 Summary and open issues

Persian has two indefinite markers, prenominal *ye(k)* and suffixed *-i*. Both forms express particular kinds of indefiniteness, as does their combination. For Modern Colloquial Persian, indefinites with *-i* express a non-uniqueness or anti-definite implication, and behave similarly to the English *any*. *Ye(k)*, on the other hand, expresses an at-issue existence implication and behaves similarly to the English *a(n)*. Finally, the combination of *ye(k)* and *NP-i* expresses an ignorance implication (Jasbi 2016). The specificity marker *-e* can be combined with *ye(k) NP* and with the com-

Klaus von Heusinger & Roya Sadeghpoor

bined form *ye(k) NP-i*, but not with (solitary) *NP-i* (Windfuhr 1979; Ghomeshi 2003). Based on these semantic functions and on the comparison of the two specificity adjectives *ein gewisser* and *ein bestimmter* in German, we hypothesized that the difference between the interpretation of the two indefinites lies in the anchoring of the indefinite either to the speaker or to some other salient discourse referent; the simple indefinite *ye(k) NP-e* is interpreted as a speaker-oriented-specific referent. The complex indefinite *ye(k) NP-e-i* is interpreted as a non-speaker-specific referent.

In two pilot acceptability tasks, we tested these two indefinites in two contexts, one that suggested a speaker-specific interpretation of the indefinite, and a second that suggested a non-speaker-specific interpretation. The first study provided some support for our hypothesis, but we also found that type of indefinite and animacy had a significant effect on interpretation. We therefore designed a second pilot study with only human indefinites. Additionally, we inserted constructions with "I heard..." and "They say...", which clearly suggest a non-speaker-oriented specificity. The results of the second study do not show a preference for the simple indefinite. However, they provide some evidence that, in neutral contexts, the simple indefinite is more acceptable with speaker orientation, and the complex with non-speaker orientation. Still, the evidence is very weak. Finally, in contexts that encode speaker ignorance ("They say...", "I heard..."), the complex indefinite was slightly more acceptable than the simple indefinite, which supports our original hypothesis.

In summary, we have seen that the complex system of indefinite marking in Modern Standard Persian provides a fruitful research environment for learning more about the formal marking of subtle semantic and pragmatic functions of noun phrases, such as specificity and the referential anchoring of nominal expressions.

Acknowledgments

References

Bisang, Walter & Kim Ngoc Quang. 2020. (In)definiteness and Vietnamese classifiers. In Kata Balogh, Anja Latrouite & Robert D. Van Valin Jr. (eds.), *Nominal anchoring: Specificity, definiteness and article systems across languages* (Topics at the Grammar-Discourse Interface). Berlin: Language Science Press.

Borik, Olga, Joan Borràs-Comes & Dària Serés. 2020. Preverbal (in)definites in Russian: An experimental study. In Kata Balogh, Anja Latrouite & Robert D. Van Valin Jr. (eds.), *Nominal anchoring: Specificity, definiteness and article systems across languages* (Topics at the Grammar-Discourse Interface). Berlin: Language Science Press.

Brustad, Kristen E. 2000. *The syntax of spoken Arabic*. Washington, D.C.: Georgetown University Press.

Dayal, Veneeta. 2017. Determining (in)definiteness in the absence of articles. In Vera Hohaus & Wanda Rothe (eds.), *Proceedings of TripleA 3. Fieldwork perspectives on the semantics of African, Asian and Austronesian languages*, 85–89. Tübingen: Universitätsbibliothek Tübingen.

Donnellan, Keith S. 1966. Reference and definite descriptions. *The Philosophical Review* 75(3). 281–304.

Dryer, Matthew S. 2005. Indefinite articles. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Ebert, Christian, Cornelia Ebert & Stefan Hinterwimmer. 2013. The interpretation of the German specificity markers *bestimmt* and *gewiss*. In Cornelia Ebert & Stefan Hinterwimmer (eds.), *Different kinds of specificity across languages* (Studies in Linguistics and Philosophy 92), 31–74. Heidelberg/Berlin: Springer.

Egli, Urs & Klaus von Heusinger. 1995. The epsilon operator and e-type pronouns. In Urs Egli, Peter Pause, Christoph Schwarze, Arnim von Stechow & Götz Wienold (eds.), *Lexical knowledge in the organization of language*, 121–141. Amsterdam/Philadelphia: John Benjamins.

Fassi-Fehri, Abdelkader. 2006. A short note on Moroccan specifics. unpublished.

Ghomeshi, Hila. 2003. Plural marking, indefiniteness and the noun phrase. *Studia Linguistica* 57(2). 47–74.

Hedberg, Nancy, Emrah Görgül & Morgan Mameni. 2009. On definiteness and specificity in Turkish and Persian. In Frédéric Mailhot (ed.), *Proceedings of the 2009 Annual conference of the Canadian Linguistics Association*, 1–15. Ottawa: Canadian Linguistic Association.

von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19(3). 245–274.

von Heusinger, Klaus. 2011. Specificity. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics. An international handbook of natural language meaning*, vol. 2 (HSK 33.2), 1025–1058. Berlin: De Gruyter.

von Heusinger, Klaus & Petra B. Schumacher. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics* 154. 117–127.

von Heusinger, Klaus. 2019. Indefiniteness and specificity. In Jeanette Gundel & Barbara Abbott (eds.), *The Oxford handbook of reference*, 146–167. Oxford: Oxford University Press.

Hincha, Georg. 1961. Beiträge zu einer Morphemlehre des Neupersischen. *Der Islam* 37(1-3). 136–201.

Jasbi, Masoud. 2014. Semantics of differential object marking in Persian. Qualification Paper, University of Stanford.

Jasbi, Masoud. 2016. Three types of indefinites in Persian: Simple, complex, and antidefinite. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Proceedings of SALT 26*, 244–263. Linguistic Society of America.

Jasbi, Masoud. 2020a. The meaning of the Persian object marker rā: What it is not, and what it (probably) is. In Richard K. Larson, Sedigheh Moradi & Vida Samiia (eds.), *Advances in Iranian linguistics*. Amsterdam: John Benjamins.

Jasbi, Masoud. 2020b. The suffix that makes Farsi nouns unique. In Richard K. Larson, Sedigheh Moradi & Vida Samiia (eds.), *Advances in Iranian linguistics*. Amsterdam: John Benjamins.

Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Oxford: Blackwell.

Karimi, Simin. 2018. Generative approaches to syntax. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *The Oxford handbook of Persian linguistics*. Oxford: Oxford University Press.

Keenan, Edward L. & Karen Ebert. 1973. A note on marking transparency and opacity. *Linguistic Inquiry* 4(3). 421–424.

Krifka, Manfred & Fereshteh Modarresi. 2016. Number neutrality and anaphoric update of pseudo-incorporated nominals in Persian (and weak definites in English). In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Proceedings of SALT 26*, 874–891. Linguistic Society of America.

Lambton, Ann K. S. 1953. *Persian grammar*. Cambridge: Cambridge University Press.

Latrouite, Anja & Robert D. Van Valin Jr. 2014. Referentiality and telicity in Lakhota and Tagalog. In Doris Gerland, Christian Horn, Anja Latrouite & Albert Ortmann (eds.), *Meaning and grammar of nouns and verbs*, 401–426. Düsseldorf: Düsseldorf University Press.

Lazard, Gilbert. 1957. *Grammaire du persan contemporain*. Paris: Klincksieck.

Lazard, Gilbert. 1992. *A grammar of contemporary Persian, English translation*. Costa Meca, CA: Mazda Publishers.

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

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language and Linguistic Theory* 19(1). 153–197.

Modarresi, Fereshteh. 2014. *Bare nouns in Persian: Interpretation, grammar and prosody*. Berlin: Humboldt-Universität zu Berlin. (Doctoral dissertation).

Modarresi, Yahya. 2018. Sociolinguistics. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *The Oxford handbook of Persian linguistics*. Oxford: Oxford University Press.

Nikravan, Pegah. 2014. Semantics and processing of weak and strong definites in colloquial Persian: Evidence from an offline questionnaire. Ottawa: Carleton University. (Master-thesis).

Onea, Edgar & Ljudmila Geist. 2011. Indefinite determiners and referential anchoring. *International Review of Pragmatics* 3(2). 194–227.

Paul, Daniel. 2008. The individuating function of the Persian "indefinite suffix". In Simin Karimi (ed.), *Aspects of Iranian linguistics*, 309–328. Cambridge: Cambridge Scholars Publishing.

Schwarz, Florian. 2013. Two types of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–558.

Šimík, Radek & Markéta Burianová. 2018. Definiteness of bare NPs as a function of clausal position: A corpus study of Czech [preprint]. In *Proceedings of FASL 26*. Michigan Slavic Publications.

Toosarvandani, Maziar & Hayedeh Nasser. 2017. Quantification in Persian. In Denis Paperno & Edward L. Keenan (eds.), *Handbook of quantifiers in natural language: Volume II*, 665–696. Berlin: Springer.

Windfuhr, Gernot L. 1979. *Persian grammar. History and state of its study*. The Hague: Mouton Publishers.

Winter, Yoad. 1997. Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy* 20(4). 399–467.

Chapter 6

Accent on nouns and its reference coding in Siwi Berber (Egypt)

Valentina Schiattarella

University of Naples, L'Orientale

The aim of this article is to investigate the position of the accent on nouns in Siwi, a Berber language spoken in the oasis of Siwa, Egypt, and to see how its alternation on the last or penultimate syllable functions in terms of reference coding. In Siwi, the role of the accent placed on nouns goes beyond the field of phonology: an analysis of original data from both spontaneous discourse and elicitations will show its functions in terms of attribution of (in)definiteness of nouns, in different environments. In order to proceed with the analysis, it is worth noting that Siwi, like all other Berber languages, does not have definite or indefinite articles.

1 Introductory remarks

Siwi is part of the Berber language family (Afro-Asiatic phylum) spoken in Morocco, Algeria, Tunisia, Libya and Egypt, as well as in Mauritania, Mali, Niger and Burkina Faso. It is the easternmost of the Berber languages, as it is the only one spoken in Egypt, in two oases: Siwa and El Gaṛa. These two oases are located in the Western Desert and are very close to the Libyan border. The main oasis, Siwa, is inhabited by over 25,000 people, including Siwi, Bedouins and Egyptians who have come from other parts of the country, and settled in the oasis mainly for work. Siwi people are almost entirely bilingual, as the vast majority of the population speaks Bedouin and/or Egyptian Arabic and Siwi.

Data for this article were collected over the course of several fieldwork trips between 2011 and 2018; collection methods mainly included recordings of spontaneous data (monologues and dialogues of variable length) and elicitation sessions with both male and female speakers, of different ages. All examples

come from my corpus, which was transcribed and translated into English with the help of my consultants. Accent will be marked with an acute accent (e. g., *á*). When the position of the accent does not emerge clearly from the recording, I refer to the transcription sessions carried out with my consultants, as they reproduce exactly what the speakers say in the original recordings.

The aim of this article is to investigate the accent alternation on nouns as a marker of reference coding. The paper will be organized as follows: after an introduction on accent in general and in some Berber languages in sections 1.1 and 1.2 respectively, and an account of previous studies on Siwi in section 1.3, I will give an overview of the notions of definiteness and indefiniteness in section 1.4. In section 2, I will discuss the accent position when the noun is isolated (section 2.1) and when it is used in discourse (sections 2.2 and 2.3). I will then establish a hierarchy between accent position and other means that the language has to convey definiteness to the noun in section 2.4, and I will conclude, in section 2.5, by presenting a construction that shows clearly how speakers use the alternation of the position of the accent to mark a distinction between definite and indefinite reference.

1.1 Some remarks on accent

Scholars usually agree on the fact that accent has the function of establishing a contrast between accented and unaccented syllables (Garde 1968: 50). Moreover, "The term *stress* is used here to refer to an abstract property of syllables within the domain of "words" (cf. Dixon and Aikhenvald 2002 for discussions of the notion *word*). A stressed syllable is likely to be pronounced with more prominence than unstressed syllables." (Goedemans & van der Hulst 2013: section 1). The position of the accent is fixed in some languages (as in Czech and French), semifixed (Latin, Polish) or free (as in Russian, Italian and English). Its position is sometimes predictable at the phonological level (for example, in light of syllable weight, the presence of non-accentuable morphemes, etc.) or at the morpho-syntactic level.

Usually, the features of the accent are: "greater loudness, higher pitch, greater duration and greater accuracy of articulation (most notably in vowels)" (Goedemans & van der Hulst 2013: section 1), but sometimes these features are not specific to accent and this makes its definition complicated. The F0 rise, for example, is also found in other prosodic phenomena, as well as vocalic length, which is linked not only to accentuation but also to the vowel's quantity.

1.2 Some remarks on accent in Berber

Scholars unanimously agree that the accent has never been properly described in Berber. This was in fact pointed out by A. Basset over sixty years ago (Basset 1952: 10). The situation has not changed much since then, even though there has been a rise in interest in this issue over the last few years, especially by virtue of the tendency to study lesser-described Berber languages, like those of the eastern part of the Berber area (Tunisia, Libya and Egypt), where the accent seems to have a more relevant role, when compared to the other Berber languages.

An overview of the accent in the domain of Berber studies was carried out by Vycichl and Chaker (1984: 103-106). In the first part of this study, Vycichl and Chaker (1984: 103-105) remark that in Guellala (Jerba, Tunisia), accent plays a peculiar role. The locative is accented on the last syllable as it is in Tamezreṭ, Tunisia: *əlmáɣrəb* 'soir' (evening), *əlmaɣrə́b* 'dans la soirée' (in the evening). Adjectives like *aməllal* 'white', distinguish a determined form *áməllal* 'the white', with the accent on the ancient definite article (the initial *a-*, according to Vycichl & Chaker 1984) while *amə́llal* means 'white' or 'a white'. The authors add that in the past, accent probably played an important role, as in the Semitic languages.

Vycichl & Chaker (1984) also noticed that with the genitive preposition *n*, the accent moves back one syllable: *agbə́n* 'Haus' (house), *elbâb n ágbən* 'die Tür des Hauses' (the door of the house). This is found with other prepositions too: *amân* 'Wasser' (water), *ɣ âman* 'ins Wasser hinein' (into the water) (Vycichl 1981: 180-181).[1] Other studies on the use of accent are those by Louali (2003) and Louali & Philippson (2004) for Siwi, Louali & Philippson (2005) for Siwi and Tuareg, and Lux & Philippson (2010) for a comparison of the accent in Tetserret and Tamasheq (Niger).

These studies show that the positioning of the accent in these Berber varieties is different from that found in Siwi, which, in contrast, shares some characteristics with other varieties spoken in Tunisia and Libya. There is a dearth of studies and oral data concerning the latter, with the exception of those produced by Brugnatelli (1986; 2005), who compared the situation found in Nafusi (Libya), where the location of the accent changes when the noun is preceded by a preposition (Beguinot 1942: 12), with the one found in Jerba (Tunisia) or Siwa (according to Vycichl & Chaker 1984):

> *uráġ*: 'fox'
> *yefkû n úraġ*: 'he gave to the fox'.

---

[1]The English translation is mine. Vycichl (1981: 180-181) uses the circumflex (ˆ) to mark accent on long vowels.

Valentina Schiattarella

This 'movement' of the accent on nouns was found by Brugnatelli in the Nafusi texts after the prepositions *n, di, in, s, ded, af, denneg* and with the exclamation particles *a/ai, ya* (Brugnatelli 1986: 64-65). The author remarks that in Beguinot's texts, the movement of the accent also takes place when the subject follows the verb (Brugnatelli 1986: 66), where other Berber languages have the annexed state. That is why the author concludes that there could be a relationship between accent and state distinction (free and annexed), which is no longer attested in Siwi and in other Berber varieties spoken in Libya (Brugnatelli 1986: 68). In Nafusi, it seems that the position of the accent is important in distinguishing two different interpretations, in the case of kinship nouns:

> *rûmmu*: 'my brother'
> *rūmmû*: 'the brother, brother' (Beguinot 1942: 28-29).

1.3 Previous studies on accent in Siwi

Accent on nouns in Siwi is insensitive to quantity as it can fall on the last or penultimate syllable of the same noun. Several authors agree that the accent on the last syllable codes indefiniteness and the accent on the penultimate syllable codes definiteness. While Louali does not recognize the function of marking locatives, previously diagnosed by Vycichl, she confirms the possibility of coding the distinction between definite and indefinite forms through accent alternation (Louali 2003: 68-69). According to Louali & Philippson (2004; 2005), the function of the accent in Siwi is morpho-syntactic because it allows the distinction of the category of the verb (accent on the first syllable of the theme) and the category of the noun (when it is isolated, it has its accent on the last syllable).

Other factors that should be taken into consideration for the prediction of the position of the accent are the presence of prepositions (*i* 'to', *s* 'with, by means of', *n* 'of', *af* 'on', *d* 'with, comitative'), which has the consequence of moving the accent one or two syllables back (Louali & Philippson 2005), and of possessive clitics, where the accent is always on the penultimate syllable. The authors also add that the position of the accent is linked to pragmatic factors (2005: 13). To conclude, Souag returns to the hypothesis formulated by Vycichl: "In general, ultimate stress marks the indefinite, penultimate the definite" (Souag 2013: 80).

This overview leads us into the discussion of the corpus of data whose analysis illustrates the position that the accent can take on the noun, in different contexts and functions. The corpus is composed of a wordlist (highlighting the position of the accent when the noun is isolated) and spontaneous texts. Even though different factors can influence the elements that are at the base of accent

formation (such as factors linked to the speaker, context, intonation and position of the word in relation to the end of the prosodic unit) I decided to use this kind of sample in order to ascertain whether the accent has functions linked to morpho-syntax and pragmatics.

The position of the accent was determined through the use of PRAAT[2] and the analysis confirms what has already been discovered by Louali and Philippson in 2004, namely that from a phonological point of view, in Siwi the accented syllable features higher pitch as its only consistent cue (e. g., Figure 1). Higher pitch on the last syllable is not linked to the fact that the noun is at the end of an intonation unit. The same goes for higher pitch on the penultimate syllable, which can be present even when a noun is at the end of an intonation unit.
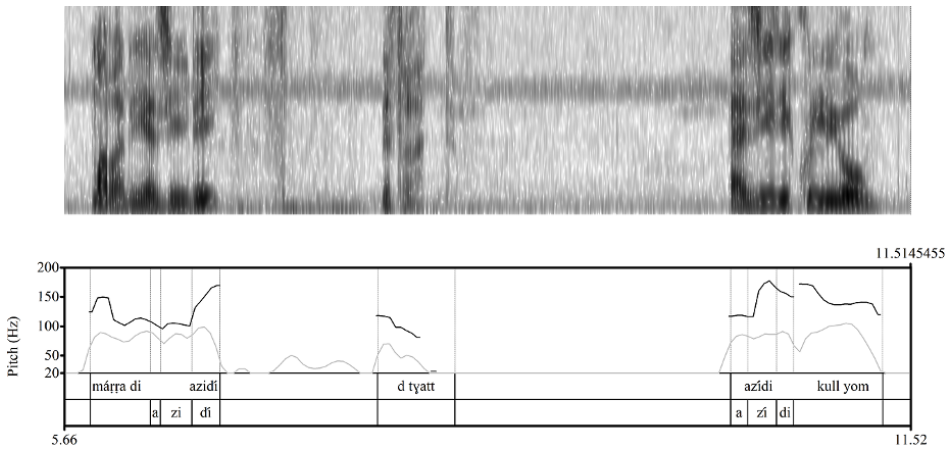
Figure 1: In this screenshot from PRAAT, the first mention of *azidi* 'jackal' shows a higher pitch on the last syllable (azi**dí**), while the second mention shows a higher pitch on the penultimate syllable (a**zí**di). Intensity is indicated by the lower and lighter line.

1.4 Definiteness and indefiniteness

There are several ways definiteness and indefiniteness can be described: authors often use terms like uniqueness, familiarity, specificity, identifiability and refer-entiality to try to explain the properties of definite nouns. While indefinites are indeed associated with the fact that the referents are generic, non-specific and non-identifiable by both hearer and speaker, definiteness is linked to concepts

---

[2]Paul Boersma & David Weenink. Praat: doing phonetics by computer [Computer program]. Version 6.1, retrieved 13 July 2019 from http://www.praat.org/

like familiarity, which refers to the possibility of a referent being recognized because it was mentioned previously (anaphora) or because it refers to the situation where it is immediately recognized by both the hearer and the speaker. Christophersen (1939: 28) adds to the concept of familiarity the feature of being based on shared knowledge between the speaker and the hearer. Referents thus do not need to be mentioned before being considered as definite. Following this, anaphora is not be interpreted only in a strict sense (the referent is definite after it is first mentioned), as the definite noun can also solely be semantically linked to a previous referent, such as 'the door' after talking about 'a house'.

Another feature of definite nouns is uniqueness, which is when there is only one possible referent the speaker could be referring to. Some nouns are more likely to be considered as definite because they are unique, such as individual nouns (*sun, Pope*, proper nouns), or nouns which are inherently relational, like some body parts and kinship terms (like *brother, leg*, etc.). Among inherently relational nouns, there are also the so-called 'functional nouns' where, in addition, the referent is unique (like *nose, mother, father*, etc.; Löbner 2011: 307). That is why in many languages, in these specific cases, definiteness is not additionally marked by a definite marker, as this could be considered redundant.

There are also some syntactic constructions that help restrict the noun in order for it to be considered definite, such as relative clauses, adnominal possessive constructions and, to a lesser extent, adjective modification. Not all languages have definite or indefinite articles (Dryer 2013a,b). Nevertheless, there is usually a way for the speaker to express whether the noun is definite or not. For example, some languages use demonstratives, which act as definite articles. Possessives can also function as definite markers.

Definiteness can be strongly determined by pragmatics in many languages, especially those without articles. Indeed, information structure and how information is conveyed is crucial and interacts with word order and with the possibility of determining whether a noun is a topic or a focus. Topic is related to what the information is about, as well as to the shared knowledge between the hearer and the speaker. In contrast, focus is "that portion of a proposition which cannot be taken for granted at the time of speech. It is the unpredictable or pragmatically non-recoverable element in an utterance. The focus is what makes an utterance into an assertion" (Lambrecht 1994: 207). That is why topic is usually associated with definiteness and focus with indefiniteness, even if this is not always the case.

A study on Polish definiteness (Czardybon 2017) has shown that topics are usually definite and in the preverbal position. When indefinites precede the verb, it is because they have to be considered to be the focus, in thetic constructions,

where the whole sentence is in focus (Lambrecht 1994: 144). The possibility of interpreting nouns as definite or indefinite is then not linked to the position (pre- or post-verbal) but by their information structure status. Information structure also interacts with other morpho-syntactic means, which determine whether a noun is definite or not.

2 The function of the accent in Siwi

The aim of this second part is to show that the possibility of coding definiteness and indefiniteness can be conveyed by accent alternation, but also to highlight how other factors can interact in giving a definite/indefinite interpretation to the noun. Definiteness is indeed conveyed through a series of factors that interact with one another: position of the accent, semantics of the noun, information structure and other morpho-syntactic elements.

2.1 The position of the accent on nouns when isolated

Nouns, when isolated, carry the accent on the last syllable and, when preceded by a preposition, they carry it on the penultimate syllable. In order to illustrate this in this first section, I have used data from elicitations, because there are other factors to consider in discourse. My data confirm those presented by Louali & Philippson (2005), who state that the accent on the noun (except for kinship terms) falls on the last syllable:

(1) aggʷíd   aggʷidán
    man.SG.M man.PL.M

(2) taltí   təltawén
    woman.SG.F woman.PL.F

However, as already noted by Vycichl (1981: 181; 2005: 207) and Louali & Philippson (2005: 12), if the noun is preceded by a preposition, the accent is on the penultimate syllable:

(3) agbə́n   g ágbən
    house.SG.M in house.SG.M

(4) amán   i áman
    water.PL.M to water.PL.M

(5) ulí   sg úli
    heart.SG.M from heart.SG.M

Valentina Schiattarella

2.2 The position of the accent in discourse: accent on the last syllable

As mentioned in section 1.3, it is usually assumed that the accent on the last syllable codes indefiniteness and the accent on the penultimate syllable codes definiteness. In this section, I will start by analyzing nouns where the accent is placed on the last syllable. In existential predicative constructions, the noun after *di* 'there is' has the accent on the last syllable, when this structure is used to introduce new referents into the discourse and they appear for the first time:

(6) di **aggʷíd**.
    EXIST man.SG.M
    'There was a man.'

(7) di **taltí** / d **aggʷíd**.
    EXIST woman.SG.F / and man.SG.M
    'There were a woman and a man.'

This is also the case for the preposition *ɣuṛ* 'at' + pronoun, when it expresses possession and the referent is generic:

(8) šal n isíwan ɣúṛ-əs **iǧəḅḅaṛə́n** dabb.
    town.SG.M of siwa at-3SG palm_tree.PL.M many
    'Siwa has a lot of palm trees.'

In the following example, both *di* and *ɣuṛ* are used to present all the main characters of the story:

(9) máṛṛa di **aggʷíd** / d::: ɣúṛ-əs **tləččá** /
    once   EXIST man.SG.M / and at-3SG girl.SG.F /

    d **akəḅḅí**. / abbá-nnəs n tlə́čča /
    and boy.SG.M / father.SG.M-POSS.3SG of girl.SG.F /
    'Once upon a time there was a man, he had a daughter and a son. The father of the girl...'

At the end of the narrations, to recapitulate the topic, Siwi uses a nonverbal predication with a pronominal demonstrative and a juxtaposed noun. This nominal predicate has the accent on the last syllable, because it is not taking up a specific referent, but is intended to recapitulate the subject of the narration:

(10) w-om **šahín**.
     DEM-2SG.F tea.SG.M
     'This is (about) the tea.'

When something is non-specific and generic, and at the same time the speaker does not need to refer to a particular category in which the referent ought to be included, we see the use of the accent on the last syllable:

(11) əssn-ím-a   ánni azídi   /
know.PFV-2PL-PRAGM COMP jackal.SG.M /

l-í-ʕəmmaṛ   ɣúṛ-əs **ankán**   / **agbə́n**   /
NEG-3SG.M-do.IPFV at-3SG place.SG.M / house.SG.M /

díma i-ʕə́mmaṛ   **taṃɣáṛt** /
always 3SG.M-do.IPFV cave.SG.F /

g **idrarə́n**.   /
in mountain.PL.M /

'Do you know that the jackal doesn't have a place, or a house, he is always in a cave, in [the] mountains.'

Later in the narration, the protagonist (a hyena) finds a cave (accent on the last syllable), and when the same cave is mentioned again, the accent is placed on the penultimate syllable. This opposition will be analyzed in more detail in 2.5.

Sometimes, a referent appears after the first mention with the accent on the last syllable: this is the case when the referent does not need to be reactivated (through anaphora), but just to be mentioned again because, for example, the speaker needs to add new information, as in the following example where the *n* + adjective construction highlights the beauty of the girl (whose birth was unexpected), even if she has just been mentioned in the previous intonation unit. The fact that it is taken up again is not intended as a strategy to mark it as known, but to qualify it:

(12) əntátət t-iráw   **tləččá**   / **tləččá**   n tkwáyəst
IDP.3SG.F 3SG.F-give_birth.PFV girl.SG.F / girl.SG.F of beautiful.SG.F
'(After she gave birth only to boys, we wanted a girl.) She gave birth to a girl'. A beautiful girl…'

If the noun is indefinite, the accent falls on the last syllable, even if it is preceded by a preposition. It is therefore worth noting here that the presence of the preposition does not obligatorily trigger the presence of the accent on the penultimate syllable, contrary to the discussion in section 2.1 with regard to nouns in isolation. In this case, the speaker is referring to a generic place:

(13) kan xsi-ṭ   aẓəṛṛá šal   s   ə́nniğ g-úni-ṭ   i
      if   want.PFV-2SG see.VN town.SG.M from up   IRR-climb.AOR-2SG to
      adrár.
      mountain.SG.M
      'If you want to see the town from the top, you will (have) to climb a
      mountain.'

(14) yə-mráq   g **ankán**   / y-ifá   aṭíl.
      3SG.M-reach.PFV in place.SG.M / 3SG.M-find.PFV garden.SG.M
      'He arrived in a place, he found a garden.'

In the following example, the author is not referring to any specific palm
tree, but he is asking the hearer to imagine a generic palm tree:

(15) ga-nə́-bdu   sg **ağəḅḅáṛ**.
      IRR-1PL-start.AOR from palm_tree.SG.M
      'Let's start from a palm tree.'

If we look at the cases listed here, we can see that nouns where the accent
is placed on the last syllable are not necessarily linked to first mention – see for
example nouns with the accent on the last syllable used when the speaker needs
to recapitulate the topic of the narration, or when an already mentioned referent
reappears, but is not crucial to the continuation of the narration. The accent on
the last syllable is then linked to the fact that the speaker needs to present the
referent, comment on it or recapitulate.

2.3 The position of the accent in discourse: accent on the penultimate syllable

In this section, examples from spontaneous discourse are presented in order to
show how placing the accent on the penultimate syllable sometimes indicates
that a noun is definite. For each example, an explanation of the kind of definite-
ness expressed will be given.

The accent on the penultimate syllable can be used with nouns mentioned
for the first time, but which are identifiable: the hearer knows who or what the
referents that the speaker is talking about are, by virtue of information provided
earlier by the speaker. In the following example, the speaker is talking about
traditions in Siwa and the hearer understands immediately that he is talking
about women from the oasis. The speaker is not referring to specific women, but
rather to a category of people:

(16) **təččíwen** tə-rṭá-ya   /
      girl.PL.F   3SG.F-cover.PFV-PRAGM /

      təltáwen   tə-rṭá-ya.   /
      woman.PL.F 3SG.F-cover.PFV-PRAGM /

      'Girls are covered, women are covered.'

The speaker is only referring to women in Siwa, and the fact of them being covered is considered as shared knowledge for both hearer and speaker. The same applies to the following example, where the sheikhs are mentioned for the first time:

(17) baʕdén yə-ʕmṛ-ín-a   / albáb   /
      then   3-do.PFV-PL-PRAGM / door.SG.M /

      i-təṃṃá-n-as   albáb   n šal.   /
      3-say.IPFV-PL-3SG.DAT door.SG.M of town.SG.M /

      ləmšáyəx yə-ʕʕə́nʕən-ən   ə́gd-əs. /
      chief.PL.M 3-sit_down.PFV-PL in-3SG /

      'Then they made a door, they call it "door of the town". The chiefs sat in it.'

Usually, when a noun appears again after a first mention it is considered to be anaphoric, but this anaphora can be also be associative: a noun is definite because it has a semantic relationship with what precedes it. In the following example, the oven is inferred from the fact that the speaker is talking about how to cook some dishes:

(18) kan **əṭṭáḅənt**   tə-ḥmá-ya
      if   oven.SG.F 3SG.F-be_hot.PFV-PRAGM
      'when the oven was hot'

In the next example, the window is mentioned for the first time, but the story is about a girl who has been kidnapped and is being held in a castle, so the presence of a window is retrievable from the situation:

(19) baʕdén tə-ẓṛ-á   sg **állon**.
      then   3SG.F-see.PFV-3SG.M.DO from window.SG.M
      'Then (Jmila) saw him from the window.'

In the following example, the well is mentioned for the first time, but it is coded as definite by virtue of it being clear that it is the only well that is present and perceivable by the characters in the castle of the sultan (visible situation use, Hawkins 1978: 110):

(20) t-uṭá   i **ánu**   n áman.
      3sg.f-fall.pfv to well.sg.m of water.pl.m
      '(The ball) fell into the well of water.'

Similarly, in the following example, the pot is mentioned for the first time, but the storyteller is asking the hearer to imagine that the woman is taking the only pot visible in the kitchen, in order to cook the chicken:

(21) t-ṛaḥ   tə-ṣṣáy   **əṭṭánǧṛət** /
      3sg.f-go.pfv 3sg.f-take.pfv pot.sg.f   /

      tə-ɣṛə́ṣ   tyaẓə́ṭ.   /
      3sg.f-slaughter.pfv chicken.sg.f /
      'She took the pot, she slaughtered a chicken.'

Proper nouns, kinship terms and toponyms, which already have a high degree of referentiality, are usually accented on the penultimate syllable, as in the following examples:

isíwan: Siwa or the people from Siwa
šáli: the citadel in the oasis of Siwa
ábba: father
wə́ltma: sister.

Placement of the accent on the penultimate syllable is therefore linked to the need to present a referent as identifiable or unique. The uniqueness of a referent can be linked both to its semantics or to its pragmatics (unique referent in the context of use). It also codes anaphora, as we will see in more detail in section 2.5.

2.4 Interaction between accent position and other strategies to mark definiteness and indefiniteness

The examples discussed in sections 2.2 and 2.3 already seem to confirm the hypothesis regarding the variation in position of the accent as a means of coding definiteness or indefiniteness. Nevertheless, in this section, I will show that the position of the accent interacts with other devices, in order to convey a definite or indefinite interpretation to the noun. These elements sometimes override the accent alternation itself. If the noun is determined by a possessive clitic, the accent is always on the penultimate syllable:

(22) **agbə́n**-nək
house.SG.M-POSS.2SG.M
'your house'

(23) **abbá**-nnəs
father.SG.M-POSS.3SG
'his father'

Adnominal possessive constructions (N + *n* 'of' + N) are definite most of the time, as the construction is a way to delimit the head noun. In the following example, the definite interpretation is conveyed by the entire construction, so the accent on the head noun can be on the last syllable:

(24) **əddhán** n **isíwan**
oil.SG.M of Siwa
'the oil of Siwa'

There are nevertheless cases where the interpretation of these constructions is indefinite, especially when they express part/whole relations, where the construction is used to refer to any generic part of the whole. In this case, both nouns have the accent on the last syllable:

(25) **ṭḥəbbə́t** n **təṃẓén**
grain.SG.F of barley.PL.F
'a grain of barley'

In general, when a noun is followed by a demonstrative, and is therefore definite, it does not always have the accent on the penultimate syllable, as one would expect. The adnominal demonstrative already codes definiteness, so most of the time, the noun has the accent on the last syllable:

(26) t-qad **tləččá tat-ók** /
3SG.M-take.PFV girl.SG.F DEM.F-2SG.M /

tə-mráq g təṭ.
3SG.F-reach.PFV in spring.SG.F
'She took this girl, she arrived at the spring.'

Cases where nouns followed by demonstratives have the accent on the penultimate syllable are nevertheless attested, especially (but not exclusively) when they appear in left-detached constructions:[3]

---

[3]This construction will be discussed in detail in 2.5.

Valentina Schiattarella

(27) i-ʕəmmáṛ-ən naknáf.   /
      3-do.IPFV-PL naknaf.SG.M /

      náknaf   daw-érwən / smiyət-ə́nnəs   /
      naknaf.SG.M DEM-2PL   / name.SG.F-POSS.3SG /

      'They prepared the *naknaf*. This *naknaf* was called… (*tqaqish*).'

When there is a preposition + N + demonstrative, the accent is on the penultimate syllable of the noun:

(28) sg **ə́lwoqt** tat-érwən
      from time.SG.F DEM.F-2PL

      'from that moment'

(29) ánni gá-ḥḥ-ax   d **iṣəṛɣénən** daw-i-(y)óm.
      COMP IRR-go.AOR-1SG and bedouin.PL.M DEM-PL-2SG.F

      'So I (can) go with these Bedouins.'

We often find the noun in right-detached constructions, which have the function of reactivating a referent (Mettouchi & Schiattarella 2018: 280), with the accent on the last syllable. In this case, the fact that the noun is in a different into-nation unit is sufficient to indicate that the referent has already been mentioned, and it needs to be reactivated. It is, then, the construction itself, not the position of the accent, that codes this function:

(30) i-lə́hhu-n   / **təṛwawén**
      3-be_happy.IPFV-PL / child.PL.F

      'They were happy, the children.'

But if a noun is inherently referential (proper nouns), the accent is on the penultimate syllable:

(31) əǧǧə́n n áddoṛ   / y-uṭə́n   / **Ḥássnin**. /
      one.M of time.SG.M / 3SG.M-get_sick.PFV / ḥassnin   /

      yə-ngə́r   yə-ṭṭís-a   g ágbən.   //
      3SG.M-stay.PFV 3SG.M-rest.PFV-PRAGM in house.SG.M //

      'Once, he got sick, Hassnin. He stayed resting at home.'

Most of the time, the referent in this right-detached construction is fol-lowed by a demonstrative, and in this case, as mentioned in reference to exam-ples (26) and (27), the position of the accent has no function in determining the definiteness of the noun.

(32) g-úɣi-x   ə́ǧǧət gə́d-sən /
     IRR-buy.AOR-1SG one.F from-3PL /

     amẓá   daw-óm.
     ogre.SG.M DEM-2SG.F

     'He thought: "I will marry one of them", this ogre.'

Contrary to descriptive relative clauses where the relative marker is not obligatory, restrictive relative clauses are introduced by *(n) wən* (SG.M/PL and sometimes SG.F) or *tən* (SG.F) (Schiattarella 2014). Head nouns in these kinds of relative clauses must usually be considered definite and have the accent on the penultimate syllable:

(33) **tálti**   wən aggʷid-ə́nnəs   yə-ṃṃút
     woman.SG.F REL   man.SG.M-POSS.3SG 3SG.M-die.PFV

     'the woman whose husband died'

This does not mean that the definite interpretation is only given by the accent, because the restrictive relative clause is already a way to restrict a head noun, giving it a definite interpretation:

(34) yə-ṭlə́b   s   ɣúṛ-əs **tləččá**
     3SG.M-ask.PFV from at-3SG girl.SG.F

     n   wən yə-xs-ét.
     of REL   3SG.M-want.PFV-3SG.F.DO

     'He asked for the girl he wanted.'

Of course, not all head nouns of restrictive relative clauses should be considered as definite, such as in the following example where the head is an indefinite pronoun:

(35) kull   **ə́ǧǧən** wən ɣúṛ-əs aṭíl
     every one.M REL   at-3SG garden.SG.M

     'everyone who has a garden'

In general, when *ə́ǧǧən* 'one' is used alone, as an indefinite pronoun, and not as a numeral, the accent is always on the penultimate syllable. When it is a numeral, the accent can also be on the last syllable.

(36) mak   **ə́ǧǧən** yə-xsá   anǧáf
     when one.M 3SG.M-want.PFV marry.VN

     'when someone wants to get married'

Valentina Schiattarella

The accent falls on the last syllable when it expresses the locative. In the introduction, I mentioned that according to Louali (2003) the locative is not expressed by the position of the accent, but this form is in fact present in our corpus. This structure is only used when the place is referential and identifiable (so it is only possible with a toponym or *ankán* 'place' + *n* 'of' + name of the place or when the name of the place is followed by a possessive). In this case, the referentiality of the noun is hierarchically more important than the fact that the accent is on the last syllable:

(37) gá-ḥḥ-aṭ **ankán** n áʕrus.
    IRR-go.AOR-2SG place.SG.M of wedding.SG.M
    'You will go to the wedding place.'

(38) yə-ččá **əlgaṛá**.
    3SG.M-eat.PFV el_gara
    'He ate in El Gara.'

(39) ənnə́gr-ax **šalí**.
    live.IPFV-1SG shali
    'I live in Shali.'

Indeed, a generic noun indicating a place cannot mark a locative solely by placing the accent on the last syllable (without the preposition):

(40) *i-nə́ddum timədrást.
    3SG.M-sleep.IPFV school.SG.F
    Intended: 'He sleeps at school.'

Locatives with the accent on the last syllable are also attested with nouns followed by a possessive. In this case, the accent is on the last syllable (which is unusual, because the accent of nouns with the possessive is always on the penultimate syllable, see examples (22) and (23)):

(41) i-təčč aksúm **timədrast-ənnə́s**.
    3SG.M-eat.IPFV meat.SG.M school.SG.F-POSS.3SG
    'He eats meat in his school.'

2.5 Same referent in close intonation units

In this section, I will analyze examples of structures where the opposition between the same noun with the accent on the last syllable and on the penultimate

164

syllable is more visible. Indeed, in many cases in the corpus analyzed here, a noun is introduced for the first time, mostly with an existential predication (*di* 'there is'), and then taken up again with the accent on the penultimate syllable, in the following intonation units:

(42) di **tləččá** tkwayə́st t-ṛáwad-as
EXIST girl.SG.F beautiful.SG.F 3SG.F-look.IPFV-3SG.DAT

sg állon.   // **tlə́čča** t-nə́ddum.
from window.SG.M // girl.SG.F 3SG.F-sleep.IPFV

'There was a beautiful girl who was watching him from the window. (Afterwards) the girl was sleeping.'

(43) g-í-ftək   **talís**
IRR-3SG.M-open.AOR tank.SG.F

g-i-sə́-ssu   amán.   /
IRR-3SG.M-CAUS-drink.AOR water.PL.M /

kan lá-di   **tális**
if   NEG-EXIST tank.SG.F

'He opens a tank and he drinks water. If there is no tank...'

(44) márṛa di   **azidí**   / d tɣátt   //
time   EXIST jackal.SG.M / and goat.SG.F //

azídi   kull yom / g-í-ffaɣ.   /
jackal.SG.M every day / IRR-3SG.M-go_out.AOR /

'Once upon a time there was a jackal and a goat. Every day, the jackal would go out.'

Left-detached constructions, used to mark a subtopic shift to what is introduced in the preceding discourse (Mettouchi & Schiattarella 2018: 278), are also characterized by this alternation. First the referent is introduced with the accent on the last syllable; it is then taken up again in a subsequent intonation unit with the accent on the penultimate syllable. In this case too, the function of the accent is anaphora:

(45) əlmanẓár **aməllál** / **amə́llal** dáw-om / w-om   **tisə́nt**. /
view.SG.M white.SG.M / white.SG.M DEM-2SG.F / DEM-2SG.F salt.SG.F /

tísənt   / ənšní n-xə́ddam-et.
salt.SG.F / IDP.1PL 1PL-work.IPFV-3SG.F.DO

'A white view, this white, it is the salt. The salt, we work it.'

Valentina Schiattarella

(46) sad_əlḥának s   **tiní**   / d   **arə́n**   /
      sad_əlḥanak with date.SG.F / and flour.PL.M /

      bass **tíni**   d   **árən**   /
      but   date.SG.F and flour.PL.M /

      l-í-ḥaṭṭu-n-**asən**   amán.
      NEG-3-put.IPFV-PL-3PL.DAT water.PL.M

      '*Sad əlḥanak* (is made) of dates and flour, but (to) the dates and flour, they
      don't add water.'

It seems that in these constructions, the placement of the accent to mark
first mention and anaphora is strictly linked to the spatial proximity of the same
referent, probably because the alternation is more easily audible when the nouns
are pronounced in a very short period of time, while it seems that other de-
vices are needed to mark the anaphoric function of a noun that has already been
mentioned, when the two instances of the noun being mentioned are far from
each other. Moreover, in the constructions discussed in this paragraph, the noun,
when taken up again, becomes the topic of the discourse, which is not always the
case when referents that have already been mentioned reappear in a discourse.

3 Discussion and conclusions

This paper has analyzed different morpho-syntactic, semantic and pragmatic fac-
tors which all contribute to the definiteness or indefiniteness of the noun, specifi-
cally when they interact with the position of the accent on the last or penultimate
syllable. It appears that the assumption that the accent on the last syllable codes
indefiniteness and the accent on the penultimate syllable codes definiteness is
too simplistic: when other factors intervene, the situation can be different. After
describing the pattern of the accent position when a noun is isolated, whether or
not it is preceded by a preposition, I described the environments where it is more
likely that the accent will be placed on the last and on the penultimate syllable.
      The accent is on the last syllable when nouns are mentioned for the first
time, especially through existential and possessive predications. Moreover, a noun
that does not need to be reactivated has the accent on the last syllable, as the
speaker is only mentioning it to allow the continuation of the narration, some-
times adding new information. The accent is on the penultimate syllable when
a noun is anaphoric and the referent is taken up a few intonation units after the
first mention. The proximity here is crucial, as the anaphoric function could also
be coded by demonstratives. The same happens with left-detached constructions

where the noun is first introduced into an intonation unit, and is then taken up again in the following intonation unit (with the accent on the penultimate syllable), and then reappears again in the form of a resumptive pronoun in the following intonation unit. Anaphora can also be associative, with the referent only semantically linked to a previous noun. The accent is also placed on the penultimate syllable when the noun is identifiable and belongs to a recognizable category for both hearer and speaker, or when it refers to something which is clearly recognizable or perceivable as unique in the particular context of use.

Some syntactic constructions allow for the restriction, and consequently the definiteness, of the head noun, namely adnominal possessive constructions and restrictive relative clauses. In the first case, the second noun of the construction usually has the accent on the penultimate syllable. Finally, some nouns are semantically referential (proper nouns, toponyms, kinship terms), so they all have the accent on the penultimate syllable.

Nevertheless, we can observe that there are some factors that interact and override the function of the accent position, when conveying (in)definiteness, such as right- and left-detached constructions or when a noun is followed by a demonstrative or a possessive, when it is a toponym, for locatives or when it is followed by a relative clause. Definiteness and indefiniteness in Siwi are thus coded in a complex way, and they are only achieved through the interaction of different elements, at different levels. The full range of aspects of interaction among all these means still needs to be studied in detail.

Further to what has been said so far, I will conclude by adding that Siwi allows all orders when only one argument (A, S or O) is present. When there are two arguments, only AVO is possible (Mettouchi & Schiattarella 2018: 288-289). Subject affixes on the verb are obligatory in Berber and the presence of a co-referent lexical noun is rare. OV is quite a rare order, as is VA, and hence most of the time nouns before the verb are subjects and nouns after the verb are objects. As nouns can be both subject and object and can have the accent on the last or penultimate syllable in preverbal or post-verbal positions, there is no relationship, synchronically, between the coding of grammatical relation and the position of the accent.

In the corpus analyzed for this study, most of the nouns in preverbal position have the accent on the penultimate syllable, while most of the nouns in post-verbal position have the accent on the last syllable. One possible explanation is that nouns in preverbal position are topics, thus conveying known information, while post-verbal nouns are focus, thus conveying unpredictable or additional information. This hypothesis still needs to be fully analyzed.

Valentina Schiattarella

Acknowledgments

I wish to thank here all the men and women consulted for this study, who accepted to collaborate with me during my stays in Siwa. I also wish to thank the editors of the volume, the reviewers, for their valuable comments and Amina Mettouchi, who gave me feedback on earlier versions of the paper.

Abbreviations

A	agent
AOR	aorist
CAUS	causative
COMP	complementizer
DAT	dative/indirect object
DEM	demonstrative
DO	direct object
EXIST	existential
F	feminine
IDP	independent pronoun
IPFV	imperfective
IRR	irrealis
M	masculine
O	object
PFV	perfective
PL	plural
POSS	possessive
PRAGM	pragmatic relevance marker
N	noun
NEG	negative
REL	relative marker
S	subject
SG	singular
V	verb
VN	verbal noun
1	first person
2	second person
3	third person
/	end of a minor prosodic unit
//	end of a major prosodic unit

References

Basset, André. 1952. *La langue berbère*. London: Oxford University Press.

Beguinot, Francesco. 1942. *Il Berbero Nefûsi di Fassâto: Grammatica, testi raccolti dalla viva voce vocabolarietti*. Rome: Istituto per l'Oriente.

Brugnatelli, Vermondo. 1986. Alternanze accentuali e morfo-sintassi nominale nel berbero orientale. In *Contributi di orientalistica, glottologia e dialettologia* (Quaderni di Acme 7), 61–72. Milano: Cisalpino-Goliardica.

Brugnatelli, Vermondo. 2005. Voyelles et accents dans l'histoire du berbère. In Pelio Fronzaroli & Paolo Marrassini (eds.), *Proceedings of the 10th meeting of Hamito-Semitic (Afroasiatic) linguistics* (Quaderni di Semitistica 25), 371–380. Firenze: Università di Firenze.

Christophersen, Paul. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Einar Munksgaard.

Czardybon, Adrian. 2017. *Definiteness in a language without articles – A study on polish* (Dissertations in Language and Cognition 3). Düsseldorf: Düsseldorf University Press.

Dryer, Matthew S. 2013a. Definite articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at `http://wals.info/chapter/37`, Accessed on 2018-11-25).

Dryer, Matthew S. 2013b. Indefinite articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at `http://wals.info/chapter/38`, Accessed on 2018-11-25).

Garde, Paul. 1968. *L'accent* (Collection SUP). Paris: Presses universitaires de France.

Goedemans, Rob & Harry van der Hulst. 2013. Fixed Stress Locations. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at `http://wals.info/chapter/14`, Accessed on 2018-11-25).

Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press.

Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.

Valentina Schiattarella

Louali, Naïma. 2003. L'accent en berbère: Catégorie grammaticale et démarcation syntaxique. In Kamal Naït-Zerrad, Rainer Voßen & Dymitr Ibriszimow (eds.), *Nouvelles études berbères. Le verbe et autres articles* (Berber Studies, Vol. 8). Köln: Rüdiger Köppe Verlag.

Louali, Naïma & Gérard Philippson. 2004. L'accent en Siwi (Berbère d'Egypte). In Bernard Bel & Isabelle Marlien (eds.), *Actes des XXVes Journées d'Étude sur la Parole (JEP 2004)*, 325–328. Fès: Association Francophone de la Communication Parlée.

Louali, Naïma & Gérard Philippson. 2005. Deux systèmes accentuels berbères: Le Siwi et le touareg. *Faits de langues* 26. 11–22.

Lux, Cécile & Gérard Philippson. 2010. L'accent en tetserret et en tamacheq: Contacts et contrastes. In Harry J. Stroomer (ed.), *Etudes Berbères v* (Berber Studies, Vol. 28), 133–164. Köln: Rüdiger Köppe Verlag.

Mettouchi, Amina & Valentina Schiattarella. 2018. The influence of state-marking on word order and information structure in Kabyle and Siwi (Berber). In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Studies in Language Companion Series*, 265–295. Amsterdam/Philadelphia: John Benjamins.

Schiattarella, Valentina. 2014. Relativization in Siwi. In Aicha Belkadi, Kakia Chatsiou & Kirsty Rowan (eds.), *Proceedings of conference on language documentation and linguistic theory 4.* London: SOAS.

Souag, Lameen. 2013. *Berber and Arabic in Siwa (Egypt): A study in linguistic contact* (Berber Studies, Vol. 37). Köln: Rüdiger Köppe Verlag.

Vycichl, Werner. 1981. Die Konstruktion des Genitivs im Berberischen der Oase Siwa. In Inge Hofmann (ed.), *Festschrift zum 60. Geburtstag von P. Anton Vorbichler*, 175–183. Vienna: Afro-Pub.

Vycichl, Werner. 2005. *Berberstudien & A Sketch of Siwi Berber (Egypt)* (Berber Studies, Vol. 10). Köln: Rüdiger Köppe Verlag.

Vycichl, Werner & Salem Chaker. 1984. Accent. In *Encyclopédie berbère*, vol. I. Abadir – Acridophagie, 103–106. Aix-en-Provence: Edisud.

Chapter 7

Indirect anaphora in a diachronic perspective: The case of Danish and Swedish

Dominika Skrzypek

Adam Mickiewicz University, Poznań

In this paper, I offer a diachronic analysis of indirect anaphora (associative anaphora), paying particular attention to the anchoring of the anaphor and the variation between definite and possessive NPs which appear in this type of bridging in Danish and Swedish between 1220 and 1550. The study is based on a corpus of authentic texts evenly distributed across languages and genres. I argue that the expression of indirect anaphora is a crucial stage in the grammaticalization of the definite article, and that the study of the spread of the incipient definite article through this context can be described in terms of strong and weak definiteness.

1 Introductory remarks

Anaphora is one of the more widely studied discourse phenomena. The term itself is derived from Greek ('carrying back', e. g., Huang 2000: 1) and is used to describe a relationship between two linguistic elements: an antecedent and an anaphor, as in the following example:

(1)　I came into <u>a spacious room</u>. *It* was sparsely decorated and rather gloomy.

The example given in (1) includes what is often considered a typical antecedent (indefNP) and a typical anaphor (a pronoun). The simplicity of the example, however, is misleading, for anaphora is a complex linguistic and cognitive phenomenon, which has duly received a great deal of attention, both within linguistic paradigms and in other fields, such as (language) philosophy, psychology,

cognitive science and artificial intelligence studies. Each is partly interested in different aspects of anaphora, and some studies subsume anaphora under a broader study of reference in discourse (e. g., Kibrik 2011). Anaphora is the central element of such theoretical proposals as Relevance Theory (Sperber & Wilson 2012) and Centering Theory (Grosz et al. 1995).

In historical linguistics, anaphora is singled out as the first stage of the grammaticalization of the definite article. What is originally a deictic element, usually a demonstrative pronoun (see Lyons 1999), begins to be used to point not only in a physical context, but also in text (anaphora).

(2)   I came into <u>a spacious room</u>. (...)
      The room was fully decorated but rather gloomy.

The use of a demonstrative to point within text involves a shift from situational to textual deixis (Lyons 1975). As the grammaticalization progresses, new uses are found for the original pronoun, as it gradually transforms into a definite article (de Mulder & Carlier 2011).

The first article-like use of the demonstrative (i. e., a use in which, in article languages, the definite article would be used) is what could more precisely be termed direct anaphora. In this type of reference the antecedent and the anaphor co-refer. A different type of anaphora is found in (3) below.

(3)   <u>My watch</u> is dead. *The battery* is flat. (after Schwarz 2000)

Even though a co-referring antecedent for the battery is lacking, the NP is definite. Definite marking (such as a definite article) is normally a signal to the hearer that the referent of the definite NP (defNP) is known, identifiable or possible to locate, and here it seems to serve the same purpose. Moreover, it is clear that the two sentences in (3) form a coherent text and the definite marking can be interpreted accordingly, in relation to another NP, namely *my watch*. The element of the preceding discourse which makes the identification of the anaphor possible will be referred to as the anchor, after Fraurud (1990) (see section 2). The relationship between *the battery* and *my watch* is anaphoric and the defNP *the battery* is an anaphor, but since the two do not co-refer, I will use the term indirect anaphor to highlight the difference between this type of relation and the direct anaphora described above. In the literature, this type of relation is also known as associative anaphora or bridging.

In this paper, I shall focus on this particular type of textual relation diachronically. In particular, I follow the typology of indirect anaphors in terms

of their type of anchoring as presented by Schwarz (2000), and address the question of the diachronic development from demonstrative pronoun through an anaphoric marker to definite article and its relation to the proposed typology of direct and indirect anaphors. For the purpose of my study I have chosen two closely related languages, Danish and Swedish, representing the eastern branch of North Germanic. I base my study on a corpus of historical texts in each language spanning 330 years, from 1220 until 1550 (see section 3). The corpus includes the oldest extant texts in each language in which there are only sporadic instances of the incipient definite article; by 1550 the article systems of both languages have reached more or less the modern form (Stroh-Wollin 2016; Skafte Jensen 2007). I am interested in particular in how indirect anaphora is expressed throughout the time of the formation of the definite article.

The aim of the paper is to fine-grain indirect anaphors and place them in a diachronic context of article grammaticalization. More specifically, I argue that not all indirect anaphors are marked as definites simultaneously, and that in this context the grammaticalizing definite article competes against two forms: bare nouns and possessives, in particular reflexive possessives.

The paper is organized as follows: I begin by defining indirect anaphora in section 2, presenting this context in detail – the aim of the section is to show how heterogeneous a context indirect anaphora is. In section 3, I present my sources and tagging principles, together with a brief overview of definiteness and its expressions in modern North Germanic languages. Section 4 presents the results, with particular focus on the forms used as indirect anaphors and on the subtypes of these anaphors. In section 5, I discuss the possible relevance of the results for the grammaticalization of the definite article. I close with conclusions and ideas for further research in section 6.

2 Indirect anaphora

Indirect anaphora has been studied mainly synchronically and in the context of definiteness; it is therefore not surprising that it has been customary to focus on defNPs as indirect anaphors. The purpose of the studies has been to establish the link between the anaphor and its anchor, or to identify the anchor. This approach is not entirely fruitful in diachronic studies. In the context of article growth, there are few examples of definite articles in the oldest texts, while many NPs are used as indirect anaphors. Although it is interesting to see in what contexts the incipient definite article may be found, this does not give us a complete picture of its grammaticalization.

Dominika Skrzypek

For the purpose of a diachronic study it is more useful to consider the context itself, irrespective of the form of the indirect anaphor. Indirect anaphora is a type of bridging reference, which, following a long tradition, I take to be a relationship between two objects or events introduced in a text or by a text, a relationship that is not spelled out and yet constitutes an essential part of the content of the text, in the sense that without this information the lack of connection between the objects or events would make the text incoherent (Asher & Lascarides 1998). This is illustrated by the following examples.

(4) I met <u>two interesting people</u> last night at a party. *The woman* was a member of Clinton's Cabinet.

(5) In the groups there was <u>one person</u> missing. *It was Mary who left*.

(6) John <u>partied</u> all night yesterday. He's going to get drunk *again* today.

(7) Jack was going to commit <u>suicide</u>. He got *a rope*.

(8) Jack locked himself out again. He had left *his keys* on the kitchen table.

(examples (4)-(7) after Asher & Lascarides 1998: 83)

It may be noted that there is a variety of expressions treated as bridging here, including, but not limited to, defNPs. In (8), it would be possible to use a defNP instead of the possessive, and most likely it would also be possible to replace the indefNP in (7) with a defNP 'the rope'. The variation in form of indirect anaphors has not been given due attention in studies thus far, while it is of fundamental importance in a diachronic study. I wish to argue for a widening of the scope of study to include other expressions, first and foremost possessive NPs (possNPs).

For indirect anaphors, although there is no antecedent, we are (mostly) able to identify some connected entity, event/activity or scenario/frame in the preceding discourse as serving a similar function ('my watch' for 'the battery'). If nominal, the 'antecedent' has been termed a *trigger* (Hawkins 1978) or an *anchor* (Fraurud 1990) for the anaphor. The two notions differ in terms of how they paint the process of referent identification. *Trigger* implies that with its articulation a number of stereotypically connected entities are activated in the hearer's mind, from among which he/she is then free to choose when the anaphor appears. Thus:

(9) We chose <u>a quiet restaurant</u>. *The menus* were modest, yet *the food* was great.

The utterance of the indefNP 'a quiet restaurant' triggers a series of connected entities, such as menus, waiters, food, other guests, cloakrooms etc. In other words, it opens up a new reference frame or reference domain (Referenzdomäne, Schwarz 2000) within which these can be found. On hearing defNPs

such as 'the waiter' or 'the table' the hearer will automatically interpret them as belonging to the restaurant mentioned earlier (though the restaurant itself may not be a familiar one, since it is presented with an indefNP). Were the speaker to choose a referent from outside this frame and mark it as definite, the hearer would probably have more trouble interpreting it correctly:

(10)   We chose <u>a quiet restaurant</u>. *The hairdresser* was rather heavy-handed and he pulled my hair with unnecessary force.

And yet, it seems unlikely that on hearing the phrase 'a quiet restaurant' the hearer automatically sees in his/her mind's eye a series of entities connected with it. In fact, were he/she to do so, it would be a very uneconomical procedure, since only some of the potential indirect anaphors will be used in the following discourse. For the most part, only some of the potential triggers become actual triggers, and when they do, only some of the wide range of possible indirect anaphors are used. Consider the following examples:

(11)   a.   Hanna hat Hans <u>erschossen</u>. *Der Knall* war bis nach Galdbach zu hören.
           'Hanna has shot Hans. The report could be heard all the way to Galdbach.'
       b.   Hanna hat Hans <u>erschossen</u>. *Die Wunde* blutet furchtbar.
           'Hanna has shot Hans. The wound is bleeding awfully.'
       c.   Hanna hat Hans <u>erschossen</u>. *Das Motiv* war Eifersucht.
           'Hanna has shot Hans. The motive was jealousy.'
       d.   Hanna hat Hans <u>erschossen</u>. Die Polizei fand *die Waffe* im Küchen-schrank.
           'Hanna has shot Hans. The police found the weapon in the kitchen cabinet.'

(Schwarz 2000: 38; she calls the collection of entities/processes activated with the use of a trigger 'konzeptueller Skopus'.)

Another term for the antecedent-like entity in preceding discourse is *anchor*, to my knowledge first introduced by Fraurud (1990). In contrast to the term *trigger*, it takes into account the actual anaphor and the process of accessing the referent by searching for an 'anchor' in the previous discourse. This term also has the value of being equally applicable to indirect and direct anaphors (the most obvious anchor would be the co-referring entity).

The examples quoted above show how heterogeneous indirect anaphora is. There are a number of relations between the anchor and the anaphor. Authors differ in their typologies of indirect anaphors; however, all of them distinguish

between at least two major types. Following Schwarz (2000) I will refer to the first type as semantic (based on lexical knowledge) and the second as conceptual (based on knowledge of the world). The former can be further subdivided into meronymic (part-whole relations) and lexical/thematic (other semantic roles), and the latter into scheme-based and inference-based. The types are illustrated with examples below.

Semantic types
a. meronymic relations

(12) A new book by Galbraith is in bookstores now. On *the cover* there is a picture of *the author*.

b. lexical/thematic relations

(13) A new book on climate change is in bookstores now. *The author* claims that mankind has only twenty years in which to make changes.

Conceptual types
a. scheme-based

(14) A charge of negligent homicide against Daw Bauk Ja could be withdrawn at the request of the *plaintiff*.

b. inference-based

(15) Wussten Sie […] dass der Schrei in Hitchcocks „Psycho" deshalb so echt wirkt, weil der Regisseur genau in dem Moment der Aufnahme eiskaltes Wasser durch *die Leitung* pumpen ließ?
      Did you know (...) that the scream in Hitchcock's *Psycho* seems so real because at the moment of filming the director allowed cold water to be pumped through the pipe?
      (example (15) from Consten 2004: 102)

To successfully interpret an anaphor of the conceptual type, a degree of knowledge of the world is necessary. The interpretation of the defNP *die Leitung* 'the pipe' relies on familiarity with the Hitchcock film and the fact that the famous scene with the scream takes place in a shower.

There are a number of other typologies of indirect anaphors (notably Irmer 2011; see also Zhao 2014 for an overview of studies of indirect anaphora), though most make similar divisions. I follow M. Schwarz's (2000) typology, since unlike the majority of other studies it is grounded in authentic texts and not constructed examples, and therefore seems best suited for a study of authentic examples, which is the subject of this paper. It should be noted, however, as Schwarz

herself frequently does, that when studying authentic texts one is often forced to classify examples that may fit more than one category, depending on what seems to be the anchor or what type of relation between the anchor and the anaphor is identified. It is also possible that in authentic texts the anaphor is accessible through more than one anchor.

Finally, a note on the form of the indirect anaphor is necessary here. Traditionally, the point of departure for all classifications has been defNPs without a co-referring antecedent. The aim of studies has been to explain their definiteness in the absence of an antecedent. However, in recent years, when the concept of bridging has become more established, more and more authors have appreciated that bridging can also occur in the absence of definites (Asher & Lascarides 1998: 107). In his discussion of totality (exhaustivity, completeness), Hawkins (1978) shows that the definite can only occur in bridging when it refers uniquely, e. g., *car − the engine* but *car − a tyre*, yet the underlying relationship between engine and car seems to be the same as that between tyre and car. It has also been demonstrated that possessives may introduce new, anchored referents (Willemse et al. 2009). Those authors found that in a considerable number of cases PM (= possessum) referents of possessive NPs are first mentions with inferential relations to the context (Willemse et al. 2009: 24). In the following, I will concentrate on the context itself and study the variety of forms found in it in historical Danish and Swedish texts.

3 Sources and tagging

The corpus used in this study consists of 29 texts in Danish and Swedish, written between 1220 and 1550, in three genres representative of the period studied: legal, religious and profane prose. From each text I chose passages with ca. 150 NPs in each (if the text was long enough), preferably high narrativity passages. The texts were divided into three periods: Period I (1220–1350), Period II (1350–1450) and Period III (1450–1550). The proposed periodization has been used in previous studies of article grammaticalization and other diachronic studies of Swedish (Delsing 2012). A total of 5822 NPs (nominal NPs only) were tagged and analyzed. The tool used for tagging and generating statistics is called DiaDef (see Figure 1 below), and was tailor-made for the project. It enables us to tag each NP for all data we assume to be in some way relevant for the choice of article, such as function in sentence (subject, object, etc.), referential status (new, unique, generic, anaphoric, etc.) and other information (case, number, gender, animacy, countability, etc.).

Dominika Skrzypek

The languages considered are both North Germanic languages of the eastern variety. The extant texts consist of Runic inscriptions from ca. 200 AD onwards; the oldest extant Danish and Swedish texts written in the Latin alphabet are legal texts from ca. 1220. For this project I look at texts from 1220 to 1550, which is a time of radical change in the grammars of both languages, including loss of case and the emergence of (in)definiteness.

Table 1: An overview of the sources

language	period	number of texts **tagged**	number of nominal NPs **extracted**
Danish	Period I (1200–1350)	7	1097
	Period II (1350–1450)	5	1016
	Period III (1450–1550)	4	787
Swedish	Period I (1200–1350)	5	1194
	Period II (1350–1450)	5	1093
	Period III (1450–1550)	3	635
		29	5822

A detailed list of quoted source texts can be found in the Sources. When quoting examples from the corpus I note the language (DA for Danish and SW for Swedish), the source text (e. g., SVT for *Sju vise mästare*; the abbreviations are also given in the Sources) and the date of its composition.

A note on the definite article in North Germanic is necessary here. The definite article is a suffix that is always attached to the noun (in the Insular Scandinavian languages Icelandic and Faroese, to the case-inflected form of the noun). Its origins are to be found in the distal demonstrative *hinn* 'yon' (e. g., Perridon 1989). Apart from the suffixed article, there are other exponents of definiteness, i. e., the weak form of the adjective (in the continental languages Danish, Swedish and Norwegian and in Faroese merely an agreement phenomenon, in Icelandic possibly retaining an original meaning of definiteness; see Naert 1969) and a preposed determiner, originally a demonstrative *sá* (in younger texts *den*) 'this'. Both the suffixed article and the preposed determiner can be combined within one NP in Swedish, Norwegian and Faroese (so-called double definiteness) but are exclusive in Danish and Icelandic. The variety of NPs is illustrated below using the example of the noun 'house' (neuter in all languages) in the singular.

(16) hus-**et** (Danish, Norwegian, Swedish)
 hús-ið (Faroese, Icelandic)
 house-DEF

(17) **det** store huset (Norwegian)
 det stora huset (Swedish)
 det store hus (Danish)
 — stóra hús-ið (Icelandic)
 hið/tað stóra hús-ið (Faroese)
 DEF large-DEF house-DEF

(18) et hus (Danish, Norwegian)
 ett hus (Swedish)
 eitt hús (Faroese)
 — hús (Icelandic)
 INDEF house

For excerption, I define bridging as widely as possible. Direct anaphora (co-reference) is tagged as DIR-A, uniques as U, generics as G, new discourse referents as NEW (when there is no connection to previous discourse whatsoever), and non-referential uses as NON-REF. For all other types of reference I use the tag INDIR-A.
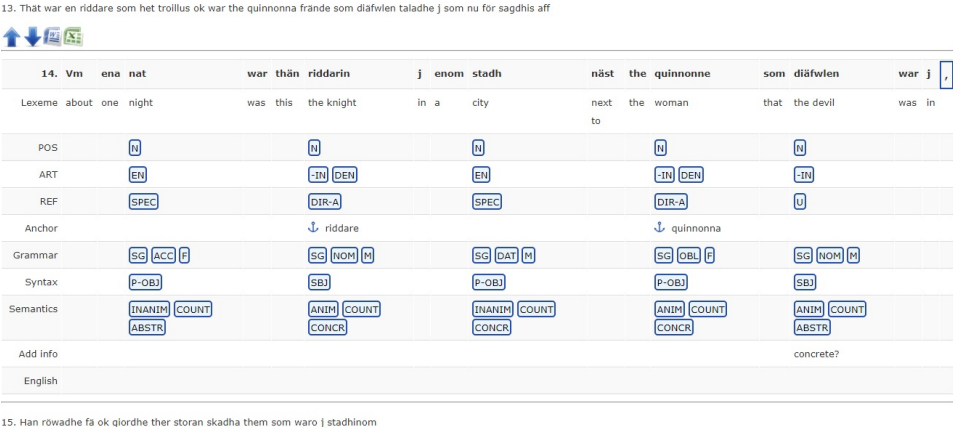
Figure 1: DiaDef print screen

The DiaDef program allows us to excerpt all NPs tagged as INDIR-A and sort them, according to the form of the NP, into: BN (bare noun), -IN (incipient definite article), POSS (possessive), DEN (demonstrative *den* 'this'), DEM (other

demonstrative elements) and EN (incipient indefinite article). For the purpose of the present study the possessives are further subdivided into POSS-GEN (genitive, e. g., *Jans* 'Jan-ɢᴇɴ'), POSS-PRO (possessive pronoun, e. g., *hans* 'his') and POSS-REFL (reflexive possessive pronoun, e. g., *sin* 'his-ʀᴇꜰʟ').

I did not expect to find large discrepancies between texts in different languages and from different periods with respect to the number of indirect anaphors in each. NPs tagged as indirect anaphors constitute ca. 25% of all NPs in the material (Table 2), with only slight variation between languages and periods. This confirms an intuitive expectation that this type of textual relation does not depend on the period. It may depend on the genre chosen; I have therefore concentrated on choosing passages of high narrativity[1] from each genre, including legal prose.

Table 2: Percentage of indirect anaphors in the corpus

period	Danish	Swedish
1200–1350	24.52%	25.80%
1350–1450	23.72%	19.67%
1450–1550	29.61%	23.62%
average	25.95%	23.03%

4 Results

I sorted all indirect anaphors according to the form of the NP. Table 3 below presents an overview of the results for each language and period.

---

[1]Old Danish and Old Swedish texts include a number of passages that can best be termed case studies, leading to the establishment of a precedent. These usually tell a short story with a number of discourse referents. I chose passages of this type over mere formulations of legal rules whenever possible.

Table 3: Indirect anaphors in Old Danish and Old Swedish according to form

	NP form	Period I 1200−1350	Period II 1350−1450	Period III 1450−1550
Danish	BN	21.85%	5.39%	4.72%
	POSS-refl	10.74%	12.45%	12.88%
	POSS-pro	26.30%	36.93%	24.46%
	POSS-gen	9.26%	12.03%	21.89%
	-IN	7.04%	10.79%	18.88%
	DEM	2.22%	1.66%	1.72%
	DEN	5.93%	6.64%	4.72%
	EN	0.74%	4.15%	0.86%
		84.08%	*90.04%*	*90.13%*
Swedish	BN	36.04%	9.30%	8.00%
	POSS-refl	4.22%	20.93%	15.33%
	POSS-pro	10.06%	22.79%	22.67%
	POSS-gen	20.78%	10.23%	20.67%
	-IN	8.77%	17.67%	11.33%
	DEM	3.90%	5.58%	6.67%
	DEN	4.22%	4.65%	6.67%
	EN	2.27%	4.65%	0.67%
		90.26%	*86.50%*	*92.01%*

First, a comment on the presentation of the results is necessary. I give percentages for each NP form used in an indirect anaphoric context; e. g., of all NPs tagged as INDIR-A in Swedish Period I, 36.04% were BNs. As can be seen from the totals (shown in italics), the forms I chose for the study cover the majority of indirect anaphors, but not all. There are other types of NPs that can be found in the material, including nouns with adjectival modifiers (adjectives in the weak or strong form) but without any other determiners. However, their frequencies were low enough for them not to be reported.

The general results show the expected patterns – a decreasing frequency of BNs in bridging reference together with a rising frequency of -IN, the incipient definite article. The high frequencies of BNs in Period I are to be expected, since in both languages the process of article grammaticalization most likely began some time before the oldest texts were written (see Skrzypek 2012: 74 for

an overview of proposed dating by different authors). The period 1220–1550 is the time when the definite article grammaticalizes in both languages. In many contexts, indirect anaphora being one of them, it comes to be used instead of BNs. We can further see that other NP types are on the rise in both languages, most notably possNPs (with reflexive possessive in Swedish and pronominal, non-reflexive possessive in Danish), not only the incipient definite article. Poss-NPs are the strongest competitor to defNPs in the material studied.

The results reported in Table 3 above show indirect anaphora without sub-dividing the context into semantic and conceptual anaphors (see section 2). They show that the context is by no means exclusively expressed by defNPs, and that possNPs in particular show high frequencies.

They also show that the major change taking place between Period I and Period II is the reduction of zero determination. In the material chosen, no BNs were found in anaphoric uses of NPs (they were still found with uniques and generics; see also Skrzypek 2012), but since the definite article is not yet fully grammaticalized it is not the default option for determination. Speakers therefore make use of other elements, most notably different types of possessives.

In the following part of the paper I will focus on the variation between defNPs and possNPs in indirect anaphora.

4.1 Semantic indirect anaphora – mereological relations

Although it may seem that I have already fine-grained the concept of indirect anaphora, the first subtype, mereological relations, is by no means homogeneous. Within it we find such different relations between anchor and anaphor as object – material (*bicycle – the steel*), object – component (*joke – the punchline*), collective – member (*deck – the card*), mass – portion (*pie – the slice*), etc. There are a number of examples of mereological relations found in the material. With limited material at my disposal, I was not able to find examples of each type of mereological relation in the Danish and Swedish texts to enable a systematic study of all sub-types for all periods in both languages. Very well represented are examples of inalienable possession, i. e., body parts, items of clothing or weaponry.

The NPs found in semantic indirect anaphora include BNs, possNPs and defNPs, although in Period I inalienables seem to be found only as BNs or poss-NPs and not as defNPs.

(19) (DA_VL 1300)

Æn of swa worthær at   man mistær allæ sinæ tændær af   *sin   høs.*
and if so be   that man loses   all   his   teeth from his.REFL head
'If it should happen so that a man loses all his teeth.'

(20) (DA_Mar 1325)

iak kom þa   fuul sørhilika til miin kæra sun ok   þahar iak sa
I   came then fully sorrowful to my   dear son and when I   saw

hanum slaa-s   mæþ næua (...) ok   spytta-s i *anlæt* ok krona-s
him   beat-PASS with fists (...) and spit-PASS in face   and crown-PASS

mæþ þorna.
with thorns

'I came full of sorrow to my dear son and as I saw he was beaten with
fists and spat in the face and crowned with thorns.'

(21) (SW_Bur 1330)

at   hon varþ   hauande mæþ guz   son ii *sino   liue*
that she became pregnant with God.GEN son in her.REFL womb

'that she carried God's son in her womb'

(22) (SW_AVL 1225)

Uærþær maþer dræpin (...) þa   skal *uighi a* þingi lysæ.
be   man   killed (...) then shall murder on ting declare

'If a man is killed then the murder shall be made public on a ting.'

In Period II, inalienables no longer appear as BNs, but either with a (reflex-
ive) possessive pronoun or the incipient definite article. It should be noted here
that North Germanic languages have retained two possessive pronouns: the reg-
ular possessive, corresponding to the English *his/her/its*, and the reflexive posses-
sive, *sin/sitt*, which is used when the possessor is the subject of the clause. The de-
fault marking of inalienables in Period II seems to be the possessive, and the incip-
ient definite article is at first only found with inalienables in direct anaphora (i. e.,
such body parts or items of clothing that are not only connected with an owner
known from previous discourse, but have also been mentioned themselves).

Dominika Skrzypek

(23)   (SW_Jart 1385)

<u>Kwinna-n</u>   gik   bort   ok   faldadhe han j   *sinom hwiff* som   <u>hon</u>
woman-DEF went away and folded   him in her   scarf   which she
hafdhe a   *sino hofdhe.*
had   on her   head
'The woman went away and folded him in her scarf which she had on her head.'

(24)   (SW_Jart 1385)

Tha syntis   quinno-n-na   hwifwir allir blodhoghir ok   water aff
then was-seen woman-GEN-DEF scarf   all   bloody   and wet   of
blodh swa at   *blodh-in*   flöt nidhir vm   quinno-n-na   kindir.
blood so   that blood-DEF flew down about woman-GEN-DEF cheeks
Hulkit herra-n   saa, ropadhe   ok   sagdhe hwar slo thik j   thit änlite
which master-DEF saw screamed and said   who   hit you in your face
älla sarghadhe. Ok quinnna-n   lypte vp *sina hand* ok   strök   sik
or   hurt   and woman-DEF lifted up her   hand and stroked herself
vm   *änliti-t* ok tha   hon tok   nidhir *hand-in-a*   tha   war
about face-DEF and when she   took down   hand-ACC-DEF then was
hon   al   blodhogh.
she (= the hand) all bloody
'Then the woman's scarf seemed all bloodied and wet with blood so that the blood flew down the woman's cheeks. Which the master saw, screamed and said "Who hit you in your face or hurt (you)?". And the woman lifted her hand and stroked her face and when she took the hand away it was all bloodied.'

Example (24) illustrates well the division of labour between the (reflexive) possessive and the incipient definite article. The possessive is used if the inalienable is mentioned for the first time (indirect anaphora). The definite article is used only in further mentions, i. e., in direct anaphora (thus *your face – the face, her hand – the hand*). Naturally, we could simply treat such examples as direct anaphors. However, it is clear that they are both co-referring with an antecedent and accessible via their anchors. It seems that this double identity, as direct and indirect anaphors, constitutes a bridging context (in the sense of Heine 2002) for defNPs to spread to indirect anaphora with meronyms. By the end of Period II and the beginning of Period III the definite article starts being used also in indirect anaphora (first mention of an inalienable possessum connected with a

known discourse referent), as shown in (25) and (26).

(25) (SW_ST 1420)

Tha bar <u>keysari-n</u>   vp *hand-ena* oc slogh hona widh
Then bore emperor-DEF up hand-DEF and hit   her   at
kinben-it   at   hon størte til iordh-inna.
cheekbone-DEF that she fell   to earth-DEF
'Then the emperor lifted his hand and hit her on the cheekbone so that
she fell down.'

(26) (DA_Jer 1480)

Tha begynthe løffwe-n som hwn war wan   gladeligh at løpe i
then began   lion-DEF as she was accustomed gladly   to run in
clostereth   (...) eller rørdhe *stiærth-en.*
monastery-DEF (...) or   wagged tail-DEF
'Then the lion began, as she was accustomed to, to gladly run in the
monastery (...) or wagged her tail.'

It should be noted that BNs are found in indirect anaphora even in Period
III; however, as illustrated in examples (27) and (28), these occurrences may be
lexicalizations rather than indirect anaphors.

(27) (DA_KM 1480)

Jamwnd-z <u>hoffui-t</u> bløde bodhe giømmen *mwn* ok *øren.*
Jamund-GEN head-DEF bled both   through mouth and ears
'Jamund's head bled both through mouth and ears.'

(28) (DA_Kat 1480)

badh   meth *mwndh oc   hiærthe.*
prayed with mouth and heart
'(She) prayed with mouth and heart.'

4.2 Semantic, lexical/thematic

The lexical/thematic type is based on our lexical knowledge of certain elements
forming more or less stereotypical events or processes, e. g., a court case involves
a judge, one or more hearings, a charge, a plaintiff and so on. In Period I we find
mostly BNs in this type of indirect anaphora (example (29)), but a few instances
of the incipient definite article have been found as well (example (30)).

(29) (SW_AVL 1225)

Sitær konæ i bo   dör *bonde.*
sits   wife in house dies husband
'If a wife is alive and the husband dies.'

(30) (SW_OgL 1280)

Nu   dræpær maþ-ær man   koma til arua man-zs-in-s   ok
now kills   man-NOM man.ACC come to heir man-GEN-DEF-GEN and

fa *drapar-a-n*   ok hugga þær niþær a fötær þæs   döþ-a.
get killer-ACC-DEF and cut   there down on feet this.GEN dead-GEN

'If a man kills another, comes to the man's heir and gets the killer and
cuts (him) down at the feet of the deceased.'

This context allows defNPs as early as Period I. I have not found possNPs in
this type of indirect anaphora. In Period II the lexical type is regularly found with
defNPs, in pairs such as *tjuven* 'the thief' – *stölden* 'the larceny', *wighia* 'ordain'
– *vixlenne* 'the ordination', *henger* 'hangs' – *galghan* 'the gallows', *rida* 'ride' –
hästen 'the horse', *fördes död* 'a dead (man) was carried' – *baren* 'the stretcher'.
Typical for this type of indirect anaphora is that the anchor need not be nominal
and the anaphor may be accessible through a VP.

4.3 Conceptual scheme-based anaphors

The conceptual types of indirect anaphora are resolved not (only) through lexical
knowledge but rather through familiarity with stereotypical relations between
objects or events and objects. The NPs found in this type are either BNs (in Period
I) or defNPs. PossNPs, on the other hand, are seldom found in this type at all,
irrespective of the period. I have located some examples of possNPs that may be
considered indirect anaphors; it should be noted that they, such as example (31),
sound natural with a reflexive possessive in Modern Swedish as well and the
choice between defNP and possNP may be a question of stylistics rather than
grammatical correctness.

(31) (SW_HML 1385)

Diäfwl-en saa hans dirue   oc reede   hanom snaru. (...) Oc baþ
devil-DEF saw his courage and prepared him   trap   (...) and asked

munk-in sik   inläta i *sin*   *cella.*
monk-DEF himself allow in his.REFL cell

'And he (the devil) asked the monk to let him in his (= the monk's) cell.'

(32) (DA_Kat 1488)

Ther sancta katherina thette fornam   tha luckthe hwn sik
when saint Catherine this   understood then locked she herself

hardeligh i *syn*   *cellæ* och badh jnderligh   till gudh.
firmly   in her.REFL cell and prayed passionately to God

'When Saint Catherine understood this, she locked herself away in her
cell and prayed passionately to God.'

However, the most commonly found NP forms in this type of indirect anaphora are either BNs (in Period I) or defNPs (sporadically in Period I, regularly in Period II and Period III), such as *tjuvnad* 'larceny' – *malseghanden* 'the plaintiff' (larceny is prosecuted, somebody sues, this person is called a plaintiff), *skuld krava* 'debt demand' – *guldit* 'the gold' (the debt is to be paid, it is possible to pay it in gold).

4.4 Conceptual inference-based

This type of indirect anaphora is the least accessible. To correctly identify the referent, the hearer must not only consider the textual information or stereotypical knowledge of the world, but also make inferences allowing him/her to resolve the anaphor. It should be noted that some authors do not consider this type anaphoric at all, e. g., Irmer (2011).

In the corpus, this type is expressed either by BNs or by defNPs. No possNPs were found here. An interesting fact, however, is that defNPs may be found as early as Period I.

(33) (SW_AVL 1225)

Maþær far sær   aþalkono gætær uiþ barn dör sv fær aþra
man   gets himself wife   begets by child dies this gets another

gætær viþ barn far hina þriðiu þör bonde þa   konæ er livændi
begets by child gets that third dies peasant than woman is alive

þa   skal af takæ hemfylgh sinæ alt þet ær   vnöt   ær hun ællær
than shall of take dowry   her all that which unused is she or

hænær börn   þa   skal hin ælsti koldær boskipti kræfiæ takær af
her   children than shall that oldest brood division demand take of

þriþiung af *bo-n-o.*
third-part of estate-DAT-DEF

'If a man marries a woman and has a child with her, after her death marries again and fathers a child and marries for the third time and dies,

Dominika Skrzypek

leaving the widow, she or her children should retrieve her dowry –all of it that is unspoilt– then the children of the first marriage demand a part in the estate and should be awarded a third of it.'

(34) (SW_Jart 1385)

Nu j the stund-in-ne for ther fram vm en prästir mz
now in this hour-DAT-DEF travelled there forward about a priest with
gud-z likama til en siukan man ok klokka-n ringde for gud-z
God-GEN body to a sick man and bell-DEF rang for God-GEN
likama.
body
'At this hour a priest was travelling to a sick man, carrying the wafer and the bell rang to announce him.'

I have not found a single example of indirect anaphora that could be classified as conceptual inference-based which would be expressed by a possNP. In this type of anaphora defNPs occur early – they are found, though only sporadically, at the beginning of Period I (while the meronymic type is not expressed with defNPs until the end of Period II). To begin with, however, BNs are prevalent. Gradually, they are suppressed by defNPs, without going through the possNP phase which the meronymic types seem to have done. This type of indirect anaphora may be seen as the one reserved for the definite article, since no other element, possessive or demonstrative, can appear here.

Table 4: NP forms of indirect anaphora in Old Danish and Old Swedish

	BN	POSS-REFL	POSS-PRO	-IN
Period I (1220–1350)	+	+	+	–
Period II (1350–1450)	–/ (+)	+	+	(+)
Period III (1450–1550)	–/ (+)	+	+	+

5 Discussion: indirect anaphora and grammaticalization of the definite article

The grammaticalization of the definite article is a relatively well-studied development, yet a number of questions remain unresolved. The first models proposed in the literature show the path from (distal) demonstrative to definite article in

188

one step (Greenberg 1978) or focus on the first stage of development, i. e., textual deixis and direct anaphora (J. Lyons 1975). Diessel (1999) sees definite articles as derived from adnominal anaphoric demonstratives, while C. Lyons (1999) argues that the origins of the definite are to be found in exophoric use (when the referent is present and accessible in the physical context) and in anaphoric use (when the referent is also easily accessible, though through discourse rather than the physical situation). Common to J. Lyons (1975), Diessel (1999) and C. Lyons (1999) is the focus on the initial stages of grammaticalization as the shift from demonstrative to definite article. However, none of these proposals account for the fact that what truly distinguishes a definite article from a demonstrative is the possibility of being used in indirect rather than direct anaphora, a context where the use of demonstratives is allowed only marginally, if at all (see Charolles 1999 for a discussion of demonstrative use in indirect anaphora). Demonstratives may, on the other hand, be used in direct anaphora without exhibiting any other properties of or grammaticalizing into definite articles. It seems therefore that the critical shift from a demonstrative to a definite article takes place where the demonstrative/incipient article appears in indirect anaphora (see also de Mulder & Carlier 2011; Skrzypek 2012).

(35)   demonstrative → direct anaphora → **indirect anaphora** → unique (→ generic)

What remains unclear is both the course of the development from direct to indirect anaphora and the course *through* indirect anaphora (which is not a homogeneous context, as demonstrated above). Also, the variation between definite article and other elements such as possessive pronouns and incipient indefinite article has not been given enough attention.

Recently, Carlier & Simonenko (2016) have proposed that the development of the definite article in French proceeds from strong to weak definiteness, with the strong-weak dichotomy, as proposed by Schwarz (2009), basically corresponding to the long-debated origins of definite meaning in either familiarity (strong definiteness) or uniqueness (weak definiteness). Based on diachronic data from Latin and French, Carlier and Simonenko suggest that the developments may be partly independent and that the weak and strong patterns unite in a single definite article with time. They note that in Classical Latin direct anaphoric relations are increasingly marked by demonstratives, among them the incipient definite article *ille*, yet the indirect anaphoric relations remain unmarked in both Classical and Late Latin and are marked with the l-article first in Old French. As Carlier and Simonenko claim, the original semantics of the l-articles involved an identity relation with a context-given antecedent (strong definiteness). With

time, an alternative definite semantics emerged, involving a presupposition of uniqueness rather than an identity relation (weak definiteness).

These two types of definiteness may be expressed by different definite articles, as has been noted for some German dialects (Austro-Bavarian German) and North Frisian (Ebert 1971), or they may correspond to different behaviours of the one definite article, as in Standard German(Schwarz 2009).

In a diachronic context, the division into strong and weak definiteness leaves indirect anaphora neither here nor there. Its resolution depends on textual anchoring (familiarity); however, it also depends on the uniqueness presupposition. Consider examples (36) and (37).

(36)   I took a taxi to the airport. The driver was a friendly elderly man.

(37)   He drove to the meeting but arrived late due to a problem with a tyre.

The use of the defNP *the driver* is based on both familiarity (with the vehicle mentioned earlier) and uniqueness (there only being one driver per car). The use of the indefNP *a tyre* is motivated by there being more than one in the given context, the anchor being the verb *drove* suggesting a vehicle, of which a tyre (the faulty tyre in this case) is a part (making the driver late). There is familiarity (we assume the existence of a vehicle) but no uniqueness. It is therefore not easy to place indirect anaphora in the strong-weak definiteness dichotomy. It may be that some types of indirect anaphora show more similarities with strong definites while others have a closer affinity with weak definites.

This would explain the relative discrepancy between inalienables and other types of indirect anaphora. The inalienable relationship between the anchor and the anaphor is based on familiarity (the anaphor being a part of the anchor) but not necessarily uniqueness. In this textual relation it is possible (and in most contexts most natural) to use the defNP *benet* 'the leg' referring to either of the two legs, just as it is to say *fickan* 'the pocket' irrespective of how many pockets there are in the outfit worn.

6 Conclusions

The model of the grammaticalization of definiteness is imperfect, as is our understanding of the category itself. It is a recurring problem in many linguistic descriptions that definites are defined mainly as text-deictic (this applies also to grammars of article-languages), whereas corpus studies show that this is not the (whole) case. While an extended deixis in the form of direct anaphora is understandable, it is by no means certain that it is the original function of the article.

Also, it is present in many languages that cannot be claimed to have definite articles, like the Slavic languages, and has not led (yet?) to the formation of a definite article. Perhaps the origins of the article are to be sought among the bridging uses, including in their widest sense (conceptual inferential).

The results of my study show that indirect anaphora is a heterogeneous context and that the incipient definite article does not spread through it uniformly in Danish and Swedish. It appears relatively early in semantic lexical types (*a book – the author*) and in conceptual types; in these contexts its main competitor is the original BNs. However, it is late in appearing in semantic meronymic types, in particular those involving inalienable possession. In this context there is strong competition from the reflexive possessive pronouns.

As indirect anaphora constitutes a crucial element of the grammaticalization of the definite article, it should be addressed in any account of the development of that article.

Acknowledgments

Sources

Danish
DA_VL = Valdemars lov, ca. 1300. Source: http://middelaldertekster.dk
DA_Jer = Af Jeronimi levned, ca. 1488, Source: Gammeldansk læsebog, 341-345
DA_Kat = Af Katherine legende, ca. 1488. Source: Gammeldansk læsebog, 346-347
DA_KM = Karl Magnus Krønike, ca. 1480. Source: Poul Lindegård Hjorth, Udg. for Universitets-Jubilæets Danske Samfund, J. H. Schultz
DA_Mar = Mariaklagen efter et runeskrevet Haandskrift-Fragment i Stockholms Kgl. Bibliotek. J. Brøndum-Nielsen & A. Rohmann (editors). 1929. Copenhagen. Manuscript Cod. Holm. A120.

Swedish – all texts downloaded from Fornsvenska textbanken.
SW_AVL = Äldre Västgötalagen, ca. 1225
SW_Bur = Codex Bureanus, ca. 1330
SW_Jart = Järteckensbok, ca. 1385

Dominika Skrzypek

SW_ST = Själens Tröst, ca. 1420
SW_OgL = Östgötalagen, ca. 1290
SW_HML = Helga manna leverne, ca. 1385

References

Asher, Nicholas & Alex Lascarides. 1998. Bridging. *Journal of Semantics* 15(1). 83–113.

Carlier, Anne & Alexandra Simonenko. 2016. The evolution of the French definite article: From strong to weak. Presentation at Going Romance.

Charolles, Michael. 1999. Associative anaphora and its interpretation. *Journal of Pragmatics* 31(3). 311–326.

Consten, Manfred. 2004. *Anaphorisch oder deiktisch? Zu einem integrativen Modell domänengebundener Referenz* (Linguistische Arbeiten, Vol. 484). Tübingen: Niemeyer.

Delsing, Lars-Olof. 2012. Stora katastrofen – med för- och efterskal. Om kasussammanfallet i fornsvenska. *Studier i svensk språkhistoria* 12. 9–29.

Diessel, Holger. 1999. *Demonstratives. Form, function and grammaticalization*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ebert, Karen. 1971. *Referenz, Sprechsituation und die bestimmten Artikel in einem nordfriesischen Dialekt* (Studien und Materialien Nr. 4.). Bredstedt: Nordfriisk Instituut.

Fraurud, Kari. 1990. Definiteness and the processing of noun phrases in natural discourse. *Journal of Semantics* 7(4). 395–433.

Greenberg, Joseph. 1978. How does a language acquire gender markers? In Joseph Greenberg (ed.), *Universals of human language 3*, 47–82. Stanford: Stanford University Press.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.

Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.

Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization*, 83–101. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Huang, Yan. 2000. *Anaphora*. Cambridge: Cambridge University Press.

Irmer, Matthias. 2011. *Bridging inferences. Constraining and resolving underspecification in discourse interpretation*. Berlin/Boston: De Gruyter.

Kibrik, Anton. 2011. *Reference in discourse*. Oxford: Oxford University Press.

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

Lyons, John. 1975. Deixis as the source of reference. In Edward L. Keenan (ed.), *Formal semantics of natural language*, 61–83. Cambridge: Cambridge University Press.

de Mulder, Walter & Anne Carlier. 2011. The grammaticalization of definite articles. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 522–534. Oxford: Oxford University Press.

Naert, Pierre. 1969. Góður maðurinn. En studie över den syntaktiska kombinationen obestämt – bestämt i isländskan, särskilt den nyare isländskan. *Arkiv för nordisk filologi* 84. 115–130.

Perridon, Harry. 1989. *Reference, definiteness and the noun phrase in Swedish*. Amsterdam: University of Amsterdam. (Doctoral dissertation).

Schwarz, Florian. 2009. *Two types of definites in natural language*. Amherst: University of Massachussets. (Doctoral dissertation).

Schwarz, Monika. 2000. *Indirekte Anaphern in Texten. Studien zur domanengebundenen Referenz und Koharenz im Deutschen*. Tübingen: Niemeyer.

Skafte Jensen, Eva. 2007. Der var engang en mand – om markeret og umarkeret artikelbrug i moderne dansk og gammeldansk. In Henrik Jørgensen & Peter Widell (eds.), *Det bedre argument. festskrift til ole togeby*, 299–320. Århus: Forlaget Wessel & Huitfeldt, Nordisk Institut.

Skrzypek, Dominika. 2012. *Grammaticalization of (in)definiteness in Swedish*. Poznań: Wydawnictwo Naukowe UAM.

Sperber, Dan & Deidre Wilson. 2012. *Meaning and relevance*. Cambridge: Cambridge University Press.

Stroh-Wollin, Ulla. 2016. The emergence of definiteness marking in Scandinavian – new answers to old questions. *Arkiv för nordisk filologi* 131. 129–169.

Willemse, Peter, Kristin Davids & Liesbet Heyvaert. 2009. English possessives as reference-point constructions and their function in the discourse. In William B. McGregor (ed.), *The expression of possession*, 13–50. Berlin: De Gruyter.

Zhao, Wei. 2014. A survey of studies of bridging anaphora. *Canadian Social Science* 10(3). 130–139.

Name index

Name Name index

Language index

Language index

Subject index

Subject index

Did you like this book?

This book was brought to you for free

language
science
press

Nominal anchoring

Set blurb on back with \BackBody{my blurb}