Linguistic Typology publishes original research on the diversity of languages and on the patterns of variation within this universe. The essence of typology lies in structural traits—ranging from sound and grammar to lexicon and discourse—that could vary independently from language to language but actually do vary together, setting limits to cross-linguistic variation and defining the groundplans on which languages are constructed. The discovery and the explanation of such interdependencies and the informed discussion of results and methods in typology are the subject matter of this journal.

Linguistic Typology invites authors to submit manuscripts to the Editor-in-chief at the address given on the title page. Instructions for contributors are available on request from the Editor-in-chief; see also the inside back cover.

Linguistic Typology is published as one volume of three issues per year (approximately 400 pages).

Subscription rates for Volume 2 (1998):
- Institutions/libraries: DM 169.00 / öS 1251.00 / sFr 162.00
- Single issues: DM 60.00 / öS 438.00 / sFr 53.00

Prices in US dollars for subscriptions in North America only:
- Institutions/libraries: US$ 100.00
- Single issues: US$ 35.00

Linguistic Typology is published by Mouton de Gruyter for the Association for Linguistic Typology (ALT). For further information on ALT and subscription details please contact the Secretary-Treasurer: Johan van der Auwera, Linguistik (GER), Universiteit Antwerpen (UAntwerp), 2610 Antwerpen, Belgium. E-mail: auwera@uiantwerp.be. ALT’s website: http://www.ling.lancs.ac.uk/alt. Institutional subscriptions can be ordered from your local bookseller or subscription agent or directly from MOUTON DE Gruyter (a division of Walter de Gruyter) at the following addresses:

For North America:
Walter de Gruyter Inc.
200 Saw Mill River Road
Hawthorne, NY 10532
USA

For all other countries:
Walter de Gruyter & Co.
Postfach 30 34 21
D-10728 Berlin
Germany

Payment may be made by credit card: MasterCard (Access), EuroCard, Visa, American Express.

A subscription order form is included at the end of the issue.

Linguistic Typology accepts paid advertising. Please address all inquiries regarding advertising to the Berlin office.

Attention before copying: Authorization to copy items for internal or personal use, or for the internal or personal use by specific clients is granted by Walter de Gruyter, for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of US$3.00 per copy is paid to CCC, 222 Rosewood Drive, Danvers, MA 01923, USA, quoting the code 1430-0532/98.


All rights reserved, including those of translation into foreign languages. No part of this journal may be reproduced in any form—by photoprint, microfilm or any other means—or transmitted or translated into machine language without written permission from the publisher. (The quotation of registered names, trade marks, etc. in this journal does not imply, even in the absence of a specific statement, that such names are exempt from laws and regulations protecting trade marks, etc., and therefore free for general use.)

Cover design by Sigurd Wendland, Berlin.

Typeset and printed in Great Britain by H. Charlesworth & Co. Ltd., Huddersfield.

Volume 2-3 (1998)

Linguistic Typology

Editor-in-chief
Frans Plank
Universität Konstanz, Germany

Editorial board
Nicholas Evans
University of Melbourne, Australia
Aleksandr E. Kibrik
Moskovskij gosudarstvennyj universitet, Russia
Marianne Mithun
University of California at Santa Barbara, USA
Edith A. Moravcsik
University of Wisconsin—Milwaukee, USA
Anna Siewierska
University of Lancaster, England
Leon Stassen
Katholieke Universiteit Nijmegen, The Netherlands
Johan van der Auwera
Universiteit Antwerpen, Belgium

Editorial assistant
Wolfgang Schellinger
Sprachwissenschaft
Universität Konstanz
D-78457 Konstanz
Germany
E-mail: frans.plank@uni-konstanz.de
Fax: +49-7531-88 4190

Editorial address
MOUTON DE GRUYTER
BERLIN • NEW YORK
Regularity in irregularity: Article use in adpositional phrases

NIKOLAUS P. HIMMELMANN

Abstract

The use of definite or specific articles in adpositional phrases often differs from that in other syntactic environments such as subject or object position. In the grammars of individual languages this is generally presented as an exception to the rules of article use in that particular language. Cross-linguistically, however, such "irregularities" are so common as to be the rule rather than the exception. Although the details of article use in adpositional phrases vary extensively even among closely related languages, a few crosslinguistic generalisations are possible. It will be shown that there are constraints on the kinds of adpositional phrases where articles can be used "irregularly", and it will be argued that these "irregularities", confined to intermediary stages of grammaticisation, need to be explained in diachronic rather than in purely semantic terms.

Keywords: adposition, agglutination, Albanian, article, case, definiteness, demonstrative, fusion, Germanic, grammaticalization, inflection, syntax, Nkore-Kiiga, relational noun, Rumanian, Tagalog

1. The problem and how it will be solved

A major problem for all theories of the definite article is to account for those instances where it is not used despite the fact that its use would be semantically well-motivated. This paper is concerned with one such context: the use or omission of articles in adpositional (i.e., pre- or postpositional) phrases. English serves as initial illustration:

(1) a. She came by bus.
   b. *She came by the bus.
   c. I take the bus.
   d. *I take bus.
In reference to a means of transportation rather than to a specific bus, the definite article is not used when *bus occurs in an adpositional phrase (1a), but it is used when *bus functions as the subject or object of a clause (1c/e). In fact, the use of the article is ungrammatical within the adpositional phrase (1b) but obligatory on subjects and objects (1d/f). Its obligatory use on subjects and objects makes it clear that the absence of an article in (1a) cannot be explained in terms of the semantics of *bus. Instead, the absence of an article in (1a) seems to be linking in some way to the fact that *bus is part of an adpositional phrase.

Examples such as (1a) are generally considered mere idiosyncrasies, not amenable to any kind of generalisation and irrelevant for theories of the article(s). As long as article use in just one language is taken into consideration, this view may seem justified. However, "irregularities" similar to those exemplified in (1) recur in language after language, provided they have article-like elements, and can therefore claim to be regular from a crosslinguistic point of view.

The crosslinguistic regularity of "irregular" article use in adpositional phrases has gone largely unnoticed in the literature. The major exception is, as in so many cases, Greenberg who very briefly and somewhat obliquely hints at this phenomenon in his seminal paper on the grammaticisation of gender marking (1978). Discussing the contexts in which articles are frequently omitted and adding a single example from Rumanian, he notes that one of these contexts "is adverbial under which may be subsumed locative and temporal constructions. These are similar to expressions in English without the article such as 'by hand', 'on foot', 'at home', and 'at night', as well as words which double as adverbs and prepositions like 'behind' and 'above'. There is here, as in other instances, a strong tendency towards grammaticalisation which takes the following form. All nouns governed by prepositions are in the non-articulated form even when their meaning is specific" (1978: 67).

Section 3 provides further empirical support for this observation. It will also be shown that the emphasis on adverbial expressions is somewhat misplaced: in several languages articles are omitted in adpositional phrases regardless of whether they function as adverbials or arguments. In these languages article use in adpositional phrases is regular insofar as it pertains to a clearly definable subclass of adpositional expressions, while still being "irregular" when compared to other syntactic environments. Since it is therefore perhaps misleading to refer to adpositional phrases in general, I will use more neutral terminology and speak of the interaction between articles and adpositions when referring to the fact that article use in adpositional phrases differs from that in other syntactic environments.

Given that the interaction between articles and adpositions is widespread, the question arises whether there are any further crosslinguistic generalisations. As will be seen, there is evidence that this interaction is indeed restricted to particular subclasses of articles and adpositions—ones which can arguably be characterised in terms of stages of their respective grammaticisation paths. These paths and the kinds of adpositions and articles (or article-like elements) which represent their different stages will be reviewed in Section 2.

Section 3 focuses on significant differences between article uses in adpositional phrases and in other syntactic environments, emphasising that it is only a certain type of adpositions (i.e., primary adpositions) which do not permit the use of articles with their complements.

Since adpositions can be further grammaticised into case markers, the question is whether these interact with articles too. In Section 4 it is proposed that the lack of articles in languages with elaborate paradigms of agglutinative case markers is another manifestation of the interaction between articles and adpositions.

Finally, in Section 5, it will be argued that the interaction between articles and adpositions cannot be explained in purely semantic or pragmatic terms and that, given the considerable crosslinguistic variability documented in Section 3, any explanation for this phenomenon has to be cast in historical terms.

The rest of this paper, thus, elaborates on these two generalisations:

— Articles are generally used less frequently and, with regard to semantic and pragmatic generalisations, less consistently in adpositional phrases than in other syntactic environments (such as subject or object position). The opposite—many semantic or pragmatic irregularities of article use in subject or object position but highly regular use in adpositional phrases—does not occur.

— The interaction between articles and adpositions is confined to the intermediary stages in their respective grammaticisation processes.

2. On the grammaticisation of adpositions and articles

2.1. Introduction

This section discusses the criteria for distinguishing kinds of articles and adpositions and outlines their grammaticisation paths. On this basis it will be possible to give a more precise statement of the central claim of
this paper, namely, that the interaction between articles and adpositions is confined to the intermediary stages in their respective grammatisations.

2.2. Adpositions

Probably the most straightforward distinction that is commonly made with regard to adpositions is that between simple adpositions, which consist of only one element (such as English on, of, in), and complex adpositions, which consist of several elements (such as English in front of, on behalf of, with respect to). The constituents of complex adpositions are fixed: they do not allow any substitutions, permutations, or inserts. Typical constituents are a simple adposition (in) and a relational noun (front). The syntactic link between these two core elements of a complex adposition and its complement is typically provided by a genitive construction (in English: of + NP). Alternatively, another simple adposition is used (e.g., to in with respect to).

In many languages it is also possible to use some relational nouns directly in adpositional function, without an additional simple adposition. Here is an example from Lewo, an Oceanic language:

(2) a-pa ŋa ra-la.
   3PL-REAL GO DUR behind-3PL.POSS
   'They went on after them.' (Early 1994: 286)

Relational nouns in this use are either body parts (head, back, etc.) or relational object parts (front, behind, centre, etc.). What distinguishes these nouns from true, more grammaticised adpositions is that they may also be used as core arguments (He hurt his back) or as the complement of a locational predicate. To illustrate the latter again from Lewo:

(3) loŋara te-ke ra-na.
   snake 3SG.BE-AT-CONT behind-3SG.POSS
   'The snake was at his rear.' (Early 1994: 97)

The complements of relational nouns are usually expressed by genitive constructions, as in these Lewo examples.

The three kinds of adpositions distinguished so far—simple adpositions, complex adpositions, and relational nouns in adpositional function—sometimes form morphosyntactically homogeneous classes. This, however, is not always the case. In particular, simple adpositions tend to be split into a variety of subclasses with substantial differences in their morphosyntactic properties. One major distinction here is between those adpositions which allow for intransitive (or adverbial) use and those which may only be used transitively (i.e., require a complement). The intransitive use of adpositions is fairly common in English (cf. put on, come in, stand out, etc.) but is impossible, for example, in Philippine or Bantu languages.

Another major distinction for simple adpositions is that between lexical (or concrete) and grammatical adpositions. Lexical adpositions (e.g., above) may co-occur with specifiers and adjuncts in addition to taking a complement. Thus, right in Bill fixed the picture right above the board near Mary’s photograph functions as a specifier of above, near Mary’s photograph as an adjunct. Grammatical adpositions do not permit specifiers and adjuncts. In English, many adpositions are found in both types of use. Example (4a) illustrates a grammatical use of on, (4b) a lexical one:

(4) a. The car crept along (*right) on three cylinders.
   b. The car crept along (right) on the main road.

Within the group of grammatical (or grammatically used) adpositions a further distinction may be made between those that are lexically governed (i.e., are part of the subcategorisation frame of a verbal expression) and those that are not. An example of a lexically governed adposition is on as used in He had to depend on himself. Lexically governed grammatical adpositions function like case markers and constitute the most grammaticalised type of adpositions.

Given these basic distinctions between different kinds of adpositions, we may now turn to their grammatisations. Figure 1 presents a schematic overview of the grammatisation path of adpositions as it is currently conceived of in the literature. The upper part of Figure 1 presents the major semantic stages in their grammatisation, ranging from concrete local to more abstract (or grammatical) meanings. The lower part outlines the major morphosyntactic stages. The framed area comprises those kinds of adpositions which interact with articles.

Of major importance here is the distinction between secondary and primary adpositions (adopted from Lehmann 1985), which cuts across between complex and simple adpositions. Secondary adpositions may be either complex or simple and are characterised by the following features: they have concrete (rather than grammatical) meanings, are polysyllabic and often also polymorphic, may be used intransitively, and tend to take genitive or oblique complements. Typical examples are English above and behind, but also Rumanian deasupra ‘over’ and afară ‘outside’. Examples of complex secondary adpositions, such as English in front of, have already been mentioned. Primary adpositions are always simple and monomorphic, cannot be used intransitively, have grammatical meanings, are mono- or bisyllabic, and tend to take nominative and accusative complements.
Figure 1. Stages in the grammaticalisation of adpositions (based on Lehmann 1985)

Obviously not all of these criteria distinguishing primary and secondary adpositions are applicable in all languages. For example, when nouns are not marked for case one cannot distinguish between oblique and accusative marked complements. But at least two features of primary adpositions—being monomorphic and permitting only transitive use—are crosslinguistically applicable and permit fairly straightforward decisions.

The two final morphosyntactic stages in the grammaticalisation of adpositions—agglutinative case markers and fusional case affixes—are not relevant here and will be taken up again in Section 4.

In general there is a close correspondence between the semantic and the morphosyntactic development of adpositions (the upper and lower parts of Figure 1). For example, in many languages concrete local relations such as superegressive and adessive are expressed by relational nouns or secondary adpositions. On the other hand, there is, to my knowledge, no language in which an accusative (the case assigned to undergoers of transitive verbs in nominative-accusative languages) is marked by a relational noun or a secondary adposition. Accusatives are typically expressed by (agglutinative or fusional) case affixes and, very rarely, also by primary adpositions (such as Rumanian pe; see Section 3.3). Still, the alignment of form and meaning in adpositional expressions does not universally follow the model suggested by Figure 1. Thus, it will occasionally be necessary to make a distinction between the morphosyntactic and the semantic aspects of the grammaticalisation of adpositions, as will become apparent shortly.

The hypothesis that only adpositions in intermediary stages of their grammaticalisation interact with articles can now be stated more precisely as follows:

- Article use in constructions involving primary adpositions or agglutinative case markers often differs from article use in other syntactic environments (such as subject or object position). No idiosyncrasies of article use are found in constructions involving relational nouns, secondary adpositions, or fusional case affixes, although these are structurally very similar and historically related to constructions involving primary adpositions or agglutinative case markers.

- Adpositions or case markers which express generalised locative meanings (such as directional, allative, ablative, locative, and benefactive markers) often constrain article use. Adpositions or case markers which express either concrete local relations (such as adessive or superegressive) or very general and abstract grammatical relations (such as accusative or genitive) do not interact with articles.

Note that the hypotheses do not predict that any particular kind of adposition will necessarily and universally interact with articles. What is predicted is that if article use in a certain type of adpositional phrase is restricted then the adposition involved will be grammaticised to an intermediary degree.

2.3. Articles

The grammaticalisation path for articles has been outlined by Greenberg in the paper on the grammaticalisation of gender markers referred to above (1978). Demonstratives may accordingly develop into definite articles (called stage I articles by Greenberg), definite articles may develop further into specific articles (also called stage II articles or non-generic articles by Greenberg), and specific articles may further develop into noun markers (Greenberg's stage III articles). The latter include gender and class markers (for further details see Greenberg 1978, 1981). As a cover term for all elements found along this grammaticalisation path, as seen in Figure 2, I use the term D-ELEMENT. This term, then, comprises demonstratives and all adnominally grammaticised elements that derive from a demonstrative (i.e., various kinds of articles and noun markers).

The major types of D-elements differ with respect to the kinds of semantic-pragmatic contexts in which they may be used. In general, the more a given element is grammaticised, the greater are the number of contexts in which it may be used. The contexts of use which we will utilise
for distinguishing demonstratives, definite articles, and specific articles are the following:\footnote{10}

- SITUATIONAL USE as in THIS GUY behind you waits to get back to his seat (referring to a person present in the utterance situation);
- DISCOURSE-DEICTIC USE as in And that's the end of THAT STORY (referring to a preceding stretch of discourse);
- TRACKING USE (also called ANAPHORIC USE) as in And a man comes along with a goat, and THE GOAT obviously is interested in the pears;
- RECOGNITION USE as in THOSE DUSTY KIND OF HILLS that they have out here by Stockton and all;
- LARGER SITUATION USE as in the sun, the Queen, the Prime Minister;
- ASSOCIATIVE-ANAPHRIC USE as in The man drove past our house in a car. THE EXHAUST FUMES were terrible;
- SPECIFIC-INDEFINITE USE as in There are THREE OTHER BOYS.

Most of these terms are taken from Hawkins's (1978, 1991) classification of the uses of the English definite article, which in turn is based on Christopherson (1939). These terms are widely used in the literature on definiteness and thus require little comment. An exception is RECOGNITION USE,\footnote{11} where the intended referent is to be identified via specific, shared knowledge rather than through situational clues or reference to preceding segments of the ongoing discourse (cf. Himmelmann 1996). Instances of such use can always be, and in fact often are, accompanied by a you know? or remember?-type of tag question. An additional example is Hit it one of THOSE BOUNCE-BACK THINGS, ... you know, the little thing that had elastic, and it has a ball.

These contexts of use may be employed as a crosslinguistically applicable diagnostic for determining the type of D-element that occurs in a given language. For the purposes of this investigation, the major types are defined in the following way; Table 1 provides a schematic overview. A DEMONSTRATIVE is a D-element which permits situational, discourse-deictic, tracking and/or recognition use, but not larger situation or associative-anaphoric use. A DEFINITIVE ARTICLE is a D-element which may be used in larger situation and associative-anaphoric contexts in addition to the four contexts in which demonstratives may occur. A SPECIFIC ARTICLE is distinguished from a definite article by allowing one more usage type, viz., specific-indefinite (typically introducing a new participant into the universe of discourse).

<table>
<thead>
<tr>
<th>DEM</th>
<th>DEF.ART</th>
<th>SPEC.ART</th>
</tr>
</thead>
<tbody>
<tr>
<td>SITUATIONAL</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>DISCOURSE-DEICTIC</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TRACKING</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RECOGNITION</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>LARGER SITUATION</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>ASSOCIATIVE-ANAPHRIC</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>SPECIFIC-INDEFINITE</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The third major kind of article commonly recognised in the literature are so-called indefinite articles. Indefinite articles derive from a different source than definite and specific articles, generally the numeral "one" (cf., for example, Givón 1981). They are thus not part of the grammaticalisation path leading from demonstratives to noun markers, and hence they are not D-elements as defined above. In most languages, the elements called "indefinite articles" do not have the same morphosyntactic properties as definite or specific articles.\footnote{12} Of particular interest here is that they never interact with adpositions (see Section 3.6). Thus, in the present paper the term ARTICLE refers exclusively to definite and specific articles.

Based on these definitions, the hypothesis that only articles in intermediary stages of their grammaticalisation interact with adpositions can be re-stated more precisely:

- If there are restrictions on the use of D-elements in adpositional phrases in a given language, these restrictions will pertain only to definite or specific articles. There will be no restrictions on the use of demonstratives or noun markers in adpositional phrases. Similarly, quantifiers, including indefinite articles, will not interact with adpositions either.

3. Crosslinguistic evidence for the "regular" interaction of articles and primary adpositions

3.1. Introduction

This section provides empirical support for the claim that only primary adpositions interact with articles. All examples have one thing in common: a nominal expression is used without an article in an adpositional phrase but with an article in other syntactic environments. Thus, nominal expressions which may be used without an article in every syntactic environment
are not included in this survey. Compare the following expressions involving school to the ones with bus in (1) above:

(5)  
  a. go to school  
  b. quit school  
  c. School starts again in September.

Unlike bus, school is used without an article in every syntactic function when referring to the institution rather than the building. Thus, the non-occurrence of an article in go to school cannot be linked to the fact that school is part of an adpositional phrase.

Furthermore, with respect to definiteness and specificity, the semantics and pragmatics of the nominal expression must be identical in the adpositional phrase and in other syntactic environments. This is important since the non-occurrence of an article within an adpositional phrase may be due to a number of factors, including semantic and pragmatic ones. Consider, for example, expressions such as English on foot or at hand, where the non-occurrence of an article is probably due to the non-referential use of the nouns. In most languages, including English, no article is used in (most) non-referential expressions, irrespective of their syntactic structure and function. Hence, on foot or at hand cannot be considered examples of the interaction between articles and adpositions.

Nonetheless, the overall generalisation established in this paper—that articles are used less frequently and less consistently in adpositional phrases than in other syntactic environments—also holds for non-referential expressions. If an article is used in a non-referential expression at all, this expression typically functions as a verbal complement rather than as the complement of an adposition (cf. play the piano vs. on piano we have the incredible George King).

With the exception of those in Section 3.4, the examples presented in the following sections all involve clearly definite uses of nominal expressions, such as the tracking use or the associative-anaphoric use. They have been carefully chosen so as to exclude any factors favouring the omission of an article other than the occurrence of the nominal expression in an adpositional phrase. In this sense, they are all straightforward examples of the interaction between articles and adpositions.

3.2. Nkore-Kiga

In many Bantu languages there are constraints on the occurrence of specific articles in prepositional phrases involving primary adpositions. No such restrictions exist for adpositional expressions formed with secondary adpositions. Furthermore, noun class markers which, according to Greenberg (1978), represent the final stage in the grammaticisation of demonstratives freely co-occur with primary adpositions. All these features are illustrated in this section with data from Nkore-Kiga (Taylor 1972, 1985).

In Nkore-Kiga, the specific article is a prefix which consists of a single, harmonic vowel. It is added to nouns in most of their uses, preceding the class prefix found in practically all uses of nouns. (6) exemplifies an indefinite (but specific) use:

(6) a-ka-cumu ku w-aa-kozesə  
    spec-cl1.2-pen cl1.2.rel(ug) 2sg-hod.past-use  
    ‘a pen, which you used (today), ...’ (Taylor 1985: 22

In some syntactic environments, however, the specific article is omitted. Taylor (1985: 88–89) mentions the following, among others:13

(i) the noun functions as the object of a negated verb;  
(ii) the noun functions as a nominal predicate;  
(iii) the noun follows a demonstrative;  
(iv) the noun is in the vocative;  
(v) the noun is the complement of a preposition.

Most relevant here is, of course, the last. There are four elements in Nkore-Kiga which may be considered primary prepositions; they are simple prepositions, are either mono- or bisyllabic, and require a complement. With two of these the specific article may co-occur, with the other two it may not (cf. Table 2).

Examele (7) illustrates the use of aha ‘on’, a preposition precluding the article:

(7) e-ki-renye ky-a-we aha ru-tara  
    spec-cl7-portion cl7-poss-2sg on cl11-shelf  
    ‘Your portion—on the shelf.’ (Taylor 1985: 82

Table 2. Interaction of articles and prepositions in Nkore-Kiga a

| +spec.art | na ‘with, and’, nka ‘like’ |
| -spec.art | aha ‘on’, onu ‘in’ |

a. In this and the following tables, a plus sign indicates that the definite or specific article may co-occur with the adpositions listed in the same row. It does not indicate that use of an article is obligatory with these adpositions. The minus sign, on the other hand, indicates that use of the article is generally excluded with the adpositions listed in the same row.
Note that the class-marking prefix ru- co-occurs with the preposition. Thus, among the two kinds of grammaticised D-elements in Nkore-Kiga, specific article and noun class markers, it is only the less strongly grammaticised specific article which interacts with primary prepositions.

As indicated in Table 2, not all primary prepositions in Nkore-Kiga interact with the specific article. No such restrictions hold for na 'with, and' and nka 'like'. When these prepositions co-occur with the specific article, the final vowel of the preposition is dropped and the preposition fuses with the prefixed article:

(8)  n-ka-ki-teera  n'-e-n-yondo
     lsg-rem.past-cl7-hit with-spec-cl9-hammer
     'I struck it with the hammer.' (Taylor 1985: 98)

A similar process is found in genitive constructions, whose genitive marker consists of a class marker (agreeing with the possessor) to which the vowel -a is suffixed. If the possessor is marked with a specific article, the final vowel of the genitive marker is dropped and the marker is fused with the expression denoting the possessor:

(9) o-mu-kira  gw'-a-ka-mi
     spec-cl3-tail cl3(poss)-spec-cl12-rabbit
     'a rabbit’s tail' (Taylor 1985: 100)

All complex prepositions involve genitive constructions, as exemplified in (10).

(10) a.  omu ma-isho  g'-e-i-hema
     in cl6-face cl6(poss)-spec-cl9-tent
     'in front of the tent' (Taylor 1985: 111)

b. aha ru-baju  oro  o-rw'-o-mu-gyera
     on cl11-side prox spec-cl11(poss)-spec-cl3-river
     'on this side of the river' (Taylor 1985: 111)

Since genitive-marked nouns permit specific articles, articles and secondary adpositions in Nkore-Kiga do not interact. (There are no simple secondary adpositions in Nkore-Kiga.)

3.3. Rumanian and Albanian

Among European languages, Rumanian and Albanian show the most pervasive and conspicuous interaction between prepositions and definite articles, again confined to primary prepositions.

Rumanian primary prepositions are distinguished from secondary prepositions by the following two features (cf. Mallinson 1986: 102, 328; Beyrer et al. 1987: 211). First, primary prepositions take their complements in the accusative case, while secondary prepositions take their complements in an oblique case (dative or genitive). Second, unlike the obligatorily transitive primary prepositions, secondary prepositions may be used intransitively (without a complement), as shown in (11):

(11)  Fata  a  rámáš  afará.
     girl.def.f.sg.nom prf.3sg stay.past.part outside
     'The girl stayed outside.' (Mallinson 1986: 328)

Only primary prepositions interact with the definite article. In construction with a primary preposition, unmodified nouns are not marked for definiteness (Beyrer et al. 1987: 91), as seen in (12):

(12) Ion  a  căzut  la  pat.
     Ion  prf.3sg fall.past.part at bed
     'Ion has taken to his bed.' (Mallinson 1986: 227)

In this example, the noun pat appears in its unmarked base form which is formally indefinite but semantically definite, as indicated by the translation.

Simple secondary prepositions allow definiteness marking, as shown by the suffix -lui which appears on the complement of the preposition deasupra in (13):

(13) Elicopter-ul  s-a  urcat
     helicopter-def.m.sg.nom rfi-prf.3sg raise.past.part deasupra munte-lui.
     over mountain-def.m.sg.gen
     'The helicopter rose over the mountain.' (Mallinson 1986: 230)

Complex secondary prepositions, consisting of a simple preposition and a noun denoting a relational object part, require their complements to be in the genitive case. As with the other secondary prepositions mentioned above, no restrictions exist with respect to definiteness marking. In (14), the complement of such a complex preposition is marked as definite:

(14) In  spate-le  gării
     in back-def.n.sg.acc station.def.f.sg.gen
     'behind the station' (Mallinson 1986: 326)

In Rumanian the relational noun in complex prepositions (spate in (14)) is always definite. Crosslinguistically this is remarkable since, as noted in Section 2.2 above, relational nouns forming part of a complex preposition are typically not marked for definiteness.15
However, the kind of preposition (primary or secondary) is not the only relevant factor. The use of an article is possible with all prepositions in Rumanian if the noun functioning as a complement is accompanied by further modifiers (e.g., an adjective or a genitive modifier). Compare (15) where the primary preposition în takes an accusative complement in both instances.

(15) a.  
\[ \text{Radu a intrat în casă.} \]
Radu Pref.3SG enter.PAST.PART in house.F.SG.ACC
‘Radu entered the house.’ (Mallinson 1986: 55)

b.  
\[ \text{Radu a intrat în casa mare.} \]
Radu Pref.3SG enter.PAST.PART in house.DEF.F.SG.ACC big.SG
‘Radu entered the big house.’ (Mallinson 1986: 55)

In (15a) no modifiers accompany the word for ‘house’, casă. Hence, it occurs in the unmarked base-form which is formally indefinite (the final vowel is a schwa). In (15b) this noun is modified by the adjective mare ‘big’, and definiteness marking becomes possible (the final a is a portmanteau morpheme, marking definiteness, gender, number, and case). Note that not every noun in a prepositional phrase which is accompanied by a modifier is automatically marked as definite. Thus, în casă mare ‘in a big house’ is also grammatical.

Furthermore, there is a minor exception to the general rule that unmodified complements cannot be marked for definiteness after primary prepositions: the two primary prepositions cu ‘with’ and pe ‘on’ do allow their complements to be marked for definiteness (cf. Mallinson 1986: 301). In the case of pe, this exception is restricted to its use as a direct object marker, as in (16a) (cf. Mallinson 1986: 88–89, 193–194; Beyrer et al. 1987: 280–281). When it is used as a locative preposition, it exhibits the same properties as all the other primary prepositions, prohibiting marking for definiteness (16b). Example (16c) illustrates the use of cu with a formally definite complement.

(16) a.  
\[ \text{O ascultam pe mama.} \]
3SG.F.ACC hear.1SG.PAST.IMPF DO mother.DEF.F.SG.ACC
‘I listened to my mother.’ (Beyrer et al. 1987: 280)

b.  
\[ \text{Ion stâr pe masă.} \]
Ion stand.3SG.PRS on table
‘Ion is standing on the table.’ (Mallinson 1986: 230)

c.  
\[ \text{apucat-ul cu mîna} \]
seize.SUP-DEF.M.SG with hand.DEF.F.SG.ACC
‘scizing with the hand’ (Mallinson 1986: 49)

As in Rumanian, the prepositions which disallow definiteness marking are primary prepositions; they assign accusative case to their complements and they may be used only transitively. Secondary prepositions assign an oblique case to their complements and allow both transitive and intransitive uses. In contrast to Rumanian, the class of primary prepositions is rather small in Albanian. According to Buchholz & Fiedler (1987: 375–384), there are altogether eleven primary prepositions, nine of which exclusively assign accusative case and two of which co-occur with both accusative and ablative complements (implying a difference in meaning). More importantly, these prepositions are not homogeneous with respect to restrictions on definiteness marking. Only five of them (listed in the top row of Table 4) interact with the definite article. For the remaining ones, which include me ‘with’, më ‘at, on’, ndër ‘under, between’, pa ‘without’, and për ‘for’, no such restrictions exist. As far as I am able to

---

**Table 3. Distribution of the definite article in prepositional phrases in Rumanian**

<table>
<thead>
<tr>
<th>+DEF.ART</th>
<th>cu ‘with’, pe when marking direct objects; prepositions taking ABL complements; all prepositions if complement is modified by an adjective, a relative clause, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>−DEF.ART</td>
<td>prepositions taking ACC-complements if complement consists of an unmodified noun</td>
</tr>
</tbody>
</table>

**Table 4. Distribution of the definite article in prepositional phrases in Albanian**

| −DEF.ART | në ‘in, at’, mbi ‘on’, nën ‘under’, përmbi ‘over’ with singular, unmodified complements; nëpër ‘through’ with singular and plural unmodified complements |
| +DEF.ART | plural and/or modified complements of the five prepositions above; all other prepositions |
ascertain, the prepositions in each of the two sets do not share any obvious (synchronic) feature other than that they either interact or do not interact with the definite article. Hence, the only generalisation possible with regard to the prepositions which impose restrictions on definiteness marking in Albanian seems to be that they are primary prepositions. But since only about half of the primary prepositions interact with the definite article it is, contrary to Rumanian, not possible to generalise across the class of primary prepositions. Instead, in Albanian the primary prepositions which interact with the definite article have to be listed individually (as has been done in Table 4).

One further difference between Albanian and Rumanian may be noted in passing. In Albanian a kind of interaction between prepositions and articles occurs which is absent from Rumanian and indeed all the other languages discussed in this paper. What we have seen so far was that the use of an article was excluded in certain kinds of adpositional phrases. In Albanian the opposite constraint occurs as well: there are two prepositions, nga 'from, out of' and te(k) 'at, to', which require the use of the definite article, unless their complements are explicitly marked as indefinite by the indefinite article or a quantifier (cf. Buchholz & Fiedler 1987: 377). These two prepositions are further distinguished from all others by the fact that they assign nominative case to their complements.

3.4. Variation among closely related languages: Germanic

In Germanic languages there are a variety of prepositional expressions in which no article occurs.16 Dialects and closely related languages vary considerably with respect to precisely which prepositional expressions do not allow the use of an article. For example, tolin hospital is preferred in British English but tolin the hospital in American English. In this section, further examples of such variation are reviewed in order to show that the non-occurrence of articles in these expressions is similar in kind to the interaction between articles and prepositions documented in the preceding sections, although it is on a much smaller scale, being restricted to fairly small sets of prepositional expressions.

Among the prepositional expressions exhibiting idiosyncrasies of article use in Germanic are the following four types: adverbial expressions such as on foot, expressions denoting an abstract concept such as in reality, expressions referring to a location and the institutional practices typically associated with that location such as in church or in town, and finally, expressions referring to a time of the day or a season. From a semantic point of view, it is debatable whether these types of expressions involve definite nominal expressions and hence whether or not the use of a definite article is to be expected on semantic grounds. At least the last two types are good candidates for larger situation use.

The possibility of a semantic explanation will be considered in Section 5 below. Here, the major concern is simply to note the substantial cross-linguistic variability. In English, as a rule, no article is used in any of the four types of prepositional expressions. In German, use of a definite article is the rule in institution-location and time expressions, but not in the two other expression types. In Swedish, use of an article is common in all expression types with the exception of the adverbial one. In Afrikaans, one may find an article even in clearly adverbial expressions. Table 5 provides some comparative data; articles are indicated by roman print.17

Table 5 oversimplifies in two respects. First, it would be wrong to infer from this data that the distribution of article use mirrors the distribution documented in Table 5 in every instance. That is, it is not always the case that an article is used in Swedish and/or Afrikaans while it may be missing in parallel expressions in English and/or German. There are several expressions in which the definite article does not occur in Afrikaans.

Table 5. Variability of article use in adpositional phrases in four Germanic languages

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>German</th>
<th>Swedish</th>
<th>Afrikaans*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ADVERBIAL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on foot</td>
<td>at home</td>
<td>zu Fuß</td>
<td>till fots</td>
<td>te voet</td>
</tr>
<tr>
<td>in reality</td>
<td>in Wirklichkeit</td>
<td>i verkligheten</td>
<td>in die werklikheid</td>
<td>by the huis'</td>
</tr>
<tr>
<td>in fashion</td>
<td>in Mode</td>
<td>på modet</td>
<td>in die mode</td>
<td></td>
</tr>
<tr>
<td>to work</td>
<td>zur Arbeit</td>
<td>till arbetet</td>
<td>(aan die) werk toe</td>
<td>in die stad</td>
</tr>
<tr>
<td>in town</td>
<td>in der Stadt</td>
<td>l sta(de)n</td>
<td>in die stad</td>
<td></td>
</tr>
<tr>
<td><strong>TIME</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in winter</td>
<td>im Winter</td>
<td>på vintern</td>
<td>in die winter</td>
<td></td>
</tr>
<tr>
<td>at dawn</td>
<td>im Morgen-</td>
<td>i gryningen</td>
<td>(met) sonop</td>
<td></td>
</tr>
</tbody>
</table>

a. Dutch is less prone than Afrikaans to use an article in these expressions. Instead, Dutch article use seems to be more similar to that of Swedish. In particular, adverbial expressions are typically articleless, cf., for example, thuis 'at home' and te voet 'on foot'.
b. This is an adverb, formally a case-inflated noun. The case inflection, however, is no longer productive but restricted to location denoting nouns in adverbial function.
c. An alternative, clearly adverbial expression is thuis.
but does occur in the other languages, for example *aan tafel* ‘at the table’ or *op kantoor* ‘at the office’ (Donaldson 1993: 62). Compare also Swedish *i norr* with English *in the north*. Second, the data in Table 5 represents the general rule for each type in the given language, but there are always exceptions. Thus, for example, while no article is used in *in Mode* ‘in fashion’ in German, *use of an article is obligatory in its antonym *aus der Mode* ‘out of fashion’ (which arguably is not lexicalised to the same degree as *in Mode*).

Furthermore, Table 5 is far from comprehensive. Many more types of prepositional expressions exist for which similar crosslinguistic variation occurs. As far as I am able to ascertain, no generalisations are possible with regard to the prepositions found in these expressions, with one exception. All types of prepositional expressions for which crosslinguistic variation occurs involve primary prepositions. Prepositional expressions based on secondary prepositions do not show any idiosyncrasies of article use.

The variation shown in Table 5 is highly relevant for the present discussion in the following way. Although, overall, the use of articles in the four Germanic languages is to a large extent identical, there are a variety of syntactic and lexical environments where it is not. The vast majority of them involve article use in prepositional expressions, as is readily apparent from the relevant sections of descriptive grammars. The variation found in prepositional phrases is specific to prepositional phrases, and no parallel variation is found in other syntactic environments. For example, article use in English and German coincides when *town* or *winter* are used in subject or object position:

\[(17) \quad \begin{align*}
    a. & \quad \text{He wanted to show the town what happened to anyone who tried to start trouble.} \\
        & \quad \text{Er wollte der Stadt zeigen, was jedem passieren würde, der} \\
        & \quad \text{versuchte, Ärger zu machen.} \\
    b. & \quad \text{Even some of the queens will die before the winter is over.} \\
        & \quad \text{Sogar einige der Königinnen sterben, bevor der Winter vorbei ist.}
\end{align*}\]

Thus, the variation observed among the Germanic languages is another instance of the general pattern investigated in this paper—that articles are, in general, used less frequently and less consistently in prepositional phrases than in other syntactic environments. This may be less readily apparent than in the earlier cases because the domain in which the interaction occurs is much less general. In Nkore-Kiga, Rumanian, and Albanian, article use was linked to an individual preposition, to a class of prepositions, or to the overall complexity of a prepositional expression.

In the Germanic languages, no generalisations are possible with respect to individual prepositions or subclasses of primary prepositions. Instead, interactions between articles and primary prepositions are confined to fairly specific types of prepositional expressions of the kind illustrated in Table 5. Still, the idiosyncrasies concern prepositional expressions. Otherwise one would expect the same kind and amount of variation in other syntactic environments. This, however, is not the case. The difference between Germanic and the languages discussed in the preceding sections, then, is not a qualitative but a quantitative one.

3.5. Tagalog

In many Western Austronesian languages, in particular in Philippine ones, the inventory of primary adpositions is very small, often consisting of only one adposition. Sometimes the definite or specific article, which is also a common feature of these languages, does not co-occur with the single primary adposition. The resulting system of noun phrase marking, in which an adposition and an article are in complementary distribution, has given rise to considerable confusion as to the nature of such a system. Taking Tagalog as an example, it will be argued in this section that the paradigm of noun phrase markers in these languages is another instance of the interaction between articles and adpositions.

In Tagalog every common noun phrase is preceded by one of the following three grammatical elements: *ang, ng* (phonetically [naŋ]), or *sa*. Proper nouns, ignored subsequently, are marked by a parallel series: *si, ni,* and *kay.*

While the functions of *ang* and, to a lesser extent, of *ng* are a matter of controversy, it is commonly agreed that *sa* is a very general locative preposition used for all kinds of arguments and adjuncts such as recipients, beneficiaries, locatives, temporal expressions. It is indeed the only primary preposition in Tagalog. (Evidence for this status includes its very broad distribution and the impossibility of an intransitive use.) It is also an obligator constituent of most complex prepositions such as *tungkōl sa* ‘about’, *para sa* ‘for’.

The marker *ang* has been analysed in a variety of ways, most commonly either as a topic marker (e.g., Schachter & Otanes 1972) or as a nominative case marker (e.g., Kroeger 1993). Counterarguments to both analyses are given in Himmelmann (1991: 8–16). The strongest argument given there for analysing *ang* as a specific article instead is the fact that several *ang*-marked phrases may co-occur in a single clause. Most importantly, in addition to the subject (or pivot) argument, predicates may also be
marked by *ang*, as in:

(18) *ang* mga buhók lamang *ang* p-in-ù-potol *ang* patalim
    SPEC PL hair only SPEC REDI-REAL(U)-cut GEN blade
    ‘(The descent of the daras on the Chinaman’s head was very gentle and) only the hair was cut by the blade.’ (Bloomfield 1917: 58)

In (18) two *ang*-marked noun phrases function as the predicate and the subject of an equative predication respectively. A more literal, structure-imitating translation would be: ‘what was being cut by the blade was only the hair’. The second *ang* is obligatory, as all common noun phrase subjects in Tagalog have to be marked by *ang*. The first *ang* is (grammatically) optional; without it, (18) would mean ‘only hair was cut by the blade’. Marking the predicate with *ang* turns it into a specific, referential expression (it’s the Chinaman’s hair which is being cut, not any other part of his body).

The use of *ang* for marking specific referential predicates cannot be explained by either the topic or the nominative analysis of *ang*. A further argument for the analysis of *ang* as a specific article is that it can be replaced by a demonstrative in all its uses, as in: 22

(19) *nag-taká* *nágá’* yung pare’
    REAL.ACT-surprise really DIST.LK priest
    ‘The priest was surprised.’

Moreover, Tagalog *ang* and the functionally equivalent markers in other Philippine languages historically derive from demonstratives (see Zorc 1977: 86; Reid 1978, 1979; Himmelmann 1997: 164). And while the grammaticising possibilities of demonstratives are very diverse, 23 they have not been shown to have ever developed into case markers.

Finally, there are a few northern Philippine languages where the element which is either cognate or at least functionally equivalent to Tagalog *ang* occurs throughout the paradigm of noun phrase markers. In Isneg, for example, the article for common noun phrases, *ya*, also co-occurs with the genitive and locative markers, as shown in Table 6 (see further Reid 1978). In the case of Isneg and similar languages, all analysts agree that the elements in question are articles.

Given, then, that *ang* is a specific article, the fact of major interest here is that it is in complementary distribution with the general locative preposition *sa*. Tagalog thus represents a language where the specific article never co-occurs with primary adpositions (of which, as mentioned above, there is only one). To illustrate, in (20) reference is made, within a prepositional expression, to a definite and specific entity (the ant is a major protagonist in the story and has been mentioned several times before) but no article is used:

(20) *na-awa’* *si* Pepito *sa* langgám
    REAL.STAT-mercy PN Pepito LOC ant
    ‘Pepito felt pity toward the ant.’

3.6. Lack of interaction between primary adpositions and demonstratives

The preceding sections provided evidence for the claim that only primary adpositions but not secondary adpositions or relational nouns interact with articles. In this section we will briefly discuss the parallel claim concerning the grammaticisation of D-elements, i.e., that it is only articles, but not demonstratives which interact with primary adpositions. The evidence is here straightforward: in all of the languages in the sample, demonstratives may be freely used with all kinds of adpositions, including primary adpositions. One example from Tagalog should suffice to illustrate this point.

(21) *dito* *sa* ka-hari’-ang itó ay merong
    PROX.LOC LOC ??-king.LOC.LK PROX PM EXIST.DIST.LOC.LK
    ísá-ng sulátan
    one.LK sultan
    ‘(and) here in this kingdom there was a sultan …’

In (21) a demonstrative (*ító*) occurs as a modifier in a prepositional phrase (*sa* ka-hari’-ang *ító*). As noted in Section 3.5, specific articles can never occur in prepositional phrases in Tagalog.

That articles and demonstratives differ in the interaction with primary adpositions should not come as a surprise. Crosslinguistically it is very common for demonstratives and articles to have different morphosyntactic properties. Dryer (1989, 1992) shows on the basis of a large and well-balanced sample that articles and nouns form a correlation pair, 25 but demonstratives and nouns do not. He notes that this result ‘presents

| Table 6. Noun phrase markers in Isneg (Barlang 1986: 94–103, 112–113) |
|-----------------------------|----------------|----------------|
| ARTICLE | GENITIVE | LOCATIVE |
| | DEF | INDEF | DEF | INDEF |
| common nouns | *ya* | *naya* | *na* | *kiya* | *kilka* |
| proper nouns | *Ce* | *ne* | *ke* |
is in line with the general hypothesis that article use is unaffected by very weakly and very strongly grammaticised adpositions, but tends to be markedly different in constructions involving adpositions in intermediary stages of their grammaticisation.

To conclude this section, another kind of interaction between articles and adpositions requires brief comment. In many languages there is a strong tendency for articles to fuse with adpositions. Well-known examples include French de le du or German in dem im. At first sight, this phenomenon looks like another instance of the kind of interaction between articles and adpositions discussed here. However, closer inspection does not show any overlap between the factors relevant for the fusion of articles and adpositions and those relevant to article use within adpositional phrases. For fusion, semantic as well as phonological factors are of primary importance, as discussed for German by Hartmann (1980), Haberland (1985), and Bisle-Müller (1991: 59–60), among others. None of these factors has been found to be relevant for the interaction of articles and adpositions. Conversely, factors of major importance for the interaction of articles and adpositions appear to be irrelevant for their fusion. In particular, articles may be fused with weakly grammaticised adpositions. German examples include fusions between an article and the combinative preposition mitm (mit dem ‘with the’) and between an article and secondary adpositions (e.g., hinterm (hinter dem) ‘behind the’). The factors contributing to the fusion of articles and adpositions indicate that the fused forms are an instance of another general tendency: the tendency of grammatical elements to fuse with each other whenever they frequently co-occur without any lexical items intervening (English examples include he'll, isn't, won't, etc.).

4. Articles and case markers

As noted in Section 2.2, the grammaticisation of adpositions does not end with primary adpositions: these may be further grammaticised into case markers. And, according to the hypothesis proposed here, the interaction between adpositions and articles is not confined to primary adpositions but also comprises the next stage on the grammaticisation path for adpositions, agglutinative case markers. No interaction occurs at the very end of this grammaticisation path, reached with fusional case affixes.

Essential characteristics of fusional case affixes are that they generally mark several categories (e.g., number and gender besides case) and that they are morphologically obligatory and often fused with the stem to which they are attached (see, e.g., Lehmann 1985). Prototypical examples are found in a variety of Indo-European languages, including Albanian and Rumanian. In both languages, definite articles are freely used with all case forms (the same holds for German and Greek, among other languages). That is, no interaction is found between fusional case affixes and articles.

Agglutinative case markers (again following Lehmann 1985) express only one category (case), they are morphologically optional (the bare stem without such a marker is a grammatical word), and the morphological boundary between stem and marker is transparent. In fact, more often than not, agglutinative case markers are phrasal clitics which appear only once per phrase, usually in phrase-final position. There is no language in our sample which displays an elaborate paradigm of agglutinative case markers. This is no chance omission. Instead, languages with elaborate paradigms of agglutinative case markers tend very strongly not to have definite or specific articles. This tendency is so strong that it may be stated as an implicational universal tendency in the sense of Comrie (1989: 19–23): If a language has an elaborate paradigm of agglutinative case markers, including markers for core cases such as nominative/ergative, accusative/absolutive, dative, and genitive as well as local cases such as locative, allative, ablative, then with overwhelmingly more than chance frequency, it will not have definite or specific articles. Languages confirming this tendency include Korean and Japanese as well as Uralic, Turkic, Dravidian, and Australian ones. The only clear exception I am presently aware of is Hungarian, a language with both an elaborate paradigm of agglutinative case markers and a definite article.

This remarkable tendency may be considered another manifestation of the interaction between articles and adpositions. When a language has an elaborate paradigm of agglutinative case markers, practically every nominal expression in this language contains an agglutinative case marker. Agglutinative case markers are relatively strongly grammaticised adpositions. Hence, it could be said that every nominal expression in such a language is a kind of adpositional phrase. Given the universal tendency for articles not to co-occur with relatively strongly grammaticised adpositions, it can be predicted that no articles are found in languages where most nominal expressions contain a relatively strongly grammaticised adposition (i.e., a primary adposition or an agglutinative case marker).

Several objections may be raised against linking the absence of articles in languages with elaborate paradigms of agglutinative case markers to the kinds of interactions between articles and adpositions reviewed in this paper. Most importantly, there are many languages which lack both articles and elaborate paradigms of agglutinative case markers (the isolating languages of east and south-east Asia, for example, as well as many Amerindian and Papuan languages). Hence the fact that a language has
an elaborate paradigm of agglutinative case markers but lacks an article could be fortuitous. However, the hypothesised link between the two phenomena may be supported by historical evidence, as will be seen in Section 5.

5. Towards an explanation

5.1. Introduction

The two major findings of the preceding sections—that the interaction between articles and adpositions is crosslinguistically widespread and that it is restricted to the intermediary stages of the grammaticalisation of D-elements and adpositions—make it highly likely that there is a principled motivation for this interaction. To help uncover it, it will first be argued in this section that an explanation in purely semantic or pragmatic terms is unsound because the details of the interaction vary considerably from language to language. Instead, the variation observed strongly suggests a diachronic explanation, allowing for gradual language-specific developments. A basic historical scenario will then be outlined which accommodates both the general tendencies and the language-particular variation. The concluding remarks will be about how and why old morphosyntax survives in synchronically fully productive constructional patterns.

5.2. The insufficiency of semantics

As mentioned in Section 3.1, there are several instances, including examples such as English on foot and at home, where the lack of an article in an adpositional phrase is semantically well-motivated. In fact, it is common practice in descriptive grammars of European languages (excluding Rumanian and Albanian) to attribute apparent idiosyncrasies in the use of articles to the semantics of these expressions, involving categories such as “adverbial” and “reference to time or institutions”, rather than recognizing that it is the adpositional nature of the expressions that might be to blame. It might be assumed accordingly that semantic factors indeed do provide the principled motivation for the widespread interaction between articles and adpositions documented here.

However, the pervasive crosslinguistic variation found in similar expressions even among closely related languages precludes an exclusively semantic explanation. In order to account for crosslinguistic variation, such an explanation would have to predict both the use and the non-use of the article in the same type of expression. For example, it might be argued that work and town in expressions such as go to work or be in town refer to generally known locations, thus being instances of the larger situation use (like the pub or the station), where a definite article is generally obligatory in all Germanic languages. Conversely, it might be argued that expressions such as to work or in town are actually non-referential expressions (similar to at home or on foot) and hence no article is to be expected. Whichever analysis is favoured, it will only account for part of the data assembled in Table 5, and it will be contradicted by the other part.

Furthermore, even within a single language a purely semantic account is often unlikely to succeed if the same kind of expression is used with an article in one syntactic environment and without in another. This was illustrated with the introductory example (1), contrasting come by bus with take the bus. It is even more obvious in the case of the non-Germanic languages discussed in Section 3, where any nominal expression is used without an article when in construction with a particular kind of adposition. And all examples from these languages did involve prototypical instances of definite use, with reference being made to uniquely identifiable entities. Hence, the syntactic environment clearly must be a major factor in determining article use in these instances.

Incidentally, this poses a (minor) problem for theories such as Hawkins’ (1978; 1991) which attempt to exclude any recourse to syntax in modelling article use. Enriching the classic Russellian account (existential quantifier plus uniqueness condition) by a set of neo-Grician implicatures, Hawkins’ theory is largely successful in accounting for actual uses of articles (in English) without recourse to syntax. However, a comprehensive theory of articles should also explain their non-use. It is primarily those semantically and pragmatically unexpected omissions of an article which appear to be syntactically determined.

Arguing against a purely semantic account of the interaction between articles and adpositions should not be misconstrued as claiming that semantics is totally irrelevant. On the contrary, in Section 3 evidence was provided that semantic factors play a role too. For example, it was found that in Nkore-Kiga, Rumanian, and Albanian comitative prepositions do not interact with articles despite the fact that they have the same morphosyntactic properties (they are obligatorily transitive and, in the case of Rumanian and Albanian, assign accusative case) as the adpositions which do interact.

The considerable crosslinguistic variation also precludes an explanation solely in syntactic terms. It would be obviously wrong to claim that primary adpositions generally interact with articles. The variation in the details of the interaction rather suggests an explanation in diachronic terms, allowing for gradual language-specific developments.
5.3. A diachronic scenario

The grammaticalisation of articles—or more precisely: the adnominal grammaticalisation of D-elements—in a given language does not proceed along identical lines in all syntactic environments. It begins in core argument (subject and object) positions, from where it may then spread to other syntactic environments. Constructions consisting of a nominal expression and secondary adpositions show no marked difference from core argument positions in the development of articles. However, in constructions involving primary adpositions there is a strong tendency for the spread to proceed much slower than in other syntactic environments. Sometimes it is completely blocked from these constructions. One way in which articles may come to be used in phrases with primary adpositions is the further grammaticalisation of secondary adpositions. If the article is regularly used with a given secondary adposition (for example, inside) and this secondary adposition is further grammaticalised into a primary adposition (in), there is no reason to discontinue the use of the article at some point along this gradual development.

This scenario is well supported by the available historical evidence for European languages. Perhaps the strongest emphasis on the gradual nature of the development is found in Behaghel's overview of the development of article use in prepositional phrases in German (1923: 75–76):


Furthermore, the scenario accommodates all the data discussed in this paper. In Tagalog, the developing specific article ang was totally blocked from spreading to adpositional phrases with the primary adposition sa. Since sa occurs in practically all adpositional phrases (including most secondary prepositions), the result of this blocking is that in Tagalog nominal expressions generally lack an article in adpositional phrases. In a similar way the article has (so far) been blocked from phrases headed by the two primary adpositions aha and omu in Nkore-Kiga.

Rumanian and Albanian represent two different stages of the further "intrusion" of the definite article into constructions with primary adpositions. In Rumanian such intrusions have to be syntactically licensed by the complexity of the nominal expression. In Albanian there is much greater variety with regard to the licensing conditions, including the complexity of the nominal expression and number marking. Most importantly, the primary adpositions are lexically subcategorised as to whether or not they impose any restrictions on the use of the definite article; about half of them don't. The Albanian data thus provide evidence for the gradual assimilation of the use of articles in adpositional phrases to that in other syntactic environments.

Germanic exemplifies a much further advanced stage in the spread of articles, with their use in adpositional phrases largely assimilated to that in other syntactic environments. It diverges only in a few minor adpositional expression types. The fact that each language exhibits a highly idiosyncratic pattern again strongly suggests that use of the article has spread in very small steps from one type of expression to the next.

Also, in this scenario the general lack of articles in languages with elaborate systems of agglutinative case markers can be seen as a manifestation of the interaction between articles and adpositions. In these languages the development of articles appears to be blocked right from the beginning because practically all nominal expressions, including those appearing in core argument positions, are in construction with relatively strongly grammaticalised adpositions.

5.4. Questions of entrenchment

Nonetheless, although it allows for crosslinguistic variability, this scenario does not provide a fully comprehensive explanation of the interaction between articles and adpositions. It raises a number of further issues, the most important of which is this: Why does the development of articles within adpositional phrases not proceed along the same lines as in other syntactic environments?

Our historical scenario is a specific adaptation of a general scenario commonly invoked for all kinds of morphological and morphosyntactic idiosyncrasies: a form or construction which is remarkable from a synchronic point of view used to be unremarkable at an earlier stage in the history of the language and thus may be explained as a retention. The general (and generally unresolved) weakness of this type of explanation is that only rarely a reason is given why a particular construction is retained. In this regard Lambrecht (1984: 758–759) notes with reference to
the related phenomenon of articleless (or bare) binomial constructions (e.g., saddle and bit, cast and crew):

I should note that there is one striking aspect of the BB [= bare binomial] phenomenon for which neither my predecessors nor I have an explanation: the very existence of these phrases. ... Traditional grammars of historical orientation, like those by Paul and by Behaghel, observe that the absence of determiners in binomials and other expressions goes back to a time in the history of the language when determiners were not yet obligatory companions of common nouns. One might then be tempted to suggest, as a possible line of explanation, that BB's are just one more example of the often observed fact that archaic language traits can be preserved in frozen form—i.e. in fixed, formulaic expressions. ... Their synchronic anomaly is tolerated, or often not even perceived, because the expressions are learned en bloc. ... But such an explanation does not address the much deeper question of why NP's that come in pairs should have preserved an archaic feature that single NP's have lost. And it begs the question of why sequences of two NP's should become formulas in the first place—in other words, why conjoined structures should be privileged candidates for formula status. But most of all, a purely historical approach cannot account for the continuing productivity of the process of BB formation.

That productivity is a crucial issue is even more palpable in the case of articleless adpositional constructions. With the exception of the Germanic examples, which could be considered fully lexicalised expressions learned en bloc, the other examples discussed here are clearly not fully lexicalised but represent the unmarked and most productive patterns for forming adpositional expressions in the respective languages.

This is not the place for an in-depth exploration of how and why fairly general and fully productive morphosyntactic patterns are retained over extended periods of time despite the “outdatedness” of their morphosyntax. Still, it would seem that Lambrecht himself advocates a position potentially providing the basic framework for solving this problem. It consists in recognizing that there is no strict and principled boundary between highly formulaic, syntactically and semantically opaque expressions and productive, semantically and syntactically fully transparent (or compositional) constructional patterns. Instead, completely invariant, lexic ally fully specified idiomatic phrases (e.g., by and large) and highly generalised and fully compositional constructional patterns (e.g., V NP) are but two extreme points on a continuum.

A specific application of this general idea is the concept of ENTRENCHMENT which plays a prominent role in Langacker’s Cognitive Grammar framework. According to Langacker (1987: 59, 380), constructional patterns or schemas are entrenched to different degrees. The more entrenched a given construction is the more its production (or activation) is automated and the more immune it is to changes occurring in other related constructions. One major factor contributing to entrenchment is frequency of use. That frequent repetition leads to entrenchment and automatisation, which in turn is closely linked to grammaticisation, has also been emphasised in recent work by Haiman (1993, 1994) and Bybee (forthcoming).

The implications of this view for explaining why the development of articles does not proceed along the same lines within adpositional phrases and in other syntactic environments should be obvious. If it can be shown that constructions consisting of a primary adpositional and a nominal expression are entrenched to a higher degree than other constructions involving nominal expressions (such as secondary adpositions and their complements or verbs and their core arguments), then it could be assumed that this difference in entrenchment provides a major reason as to why changes like those affecting the structure of nominal expressions in less entrenched syntactic environments do not occur, or occur only with some delay, in adpositional expressions involving primary adpositions.

To be workable, this account requires diagnostics for high entrenchment and lack of full compositionalitiy. In the case of productive but still clearly formulaic patterns such as binomials there indeed are numerous formal and semantic constraints which show that the construction is processed, at least in part, as an entrenched unit—such as idiosyncratic case marking and agreement patterns, the inability to occur with indefinite determiners, and the presupposition of a common semantic frame (as mentioned by Lambrecht 1984).

But what about much more general and productive constructional patterns such as phrases consisting of a primary adpositional and a nominal expression? Do these constructions also show high entrenchment and non-compositional features, thus differing from similarly productive, but less entrenched constructional patterns? Elsewhere (Himmelmann 1997: 134–157) I have suggested an affirmative answer for constructions involving grammatical elements such as primary adpositions, articles, auxiliaries, and possibly also complementisers—i.e., grammatical elements which have been grammaticalised to an intermediary degree. As diagnostics for relatively high entrenchment and the lack of full compositionality of these constructions the following two criteria have been proposed.37

First, syntactically, an otherwise highly general and productive constructional pattern lacks full compositionality if it is unclear which constituent is the head of the construction. That is, the features which are commonly attributed to syntactic heads (such as being obligatory and being the category determinant) are not assigned to a single core
constituent, but are distributed among the two core constituents, the grammaticalised construction marker (for example, the article in article-noun constructions) and the semantic head of the construction (the noun in article-noun constructions). The controversial constructions in the recently re-opened and still open-ended debate about headedness all involve elements which are grammaticalised to an intermediary degree—i.e., articles, auxiliaries, and complementisers. If the syntactic make-up of these constructions were fully transparent, there would be no need for, and no substance to, such an extended debate.

Second, semantically, an otherwise highly general and productive constructional pattern lacks full compositionality if its use is not completely predictable on semantic (and pragmatic) grounds. One way in which this lack of semantic compositionality becomes manifest is the occasional use of a constructional pattern without being semantically or pragmatically warranted. (For example, definite articles with proper nouns are semantically a case of overdetermination.)

This hypothesis and the associated diagnostics are still in need of further elaboration and testing. The concern in Himmelmann (1997) is exclusively with constructions involving articles (adnominally grammaticalised D-elements) for which the two criteria just mentioned are discussed at length. As for constructions involving primary adpositions, the following preliminary observations suggest that they are also highly entrenched and lack full compositionality.

As for syntactic transparency, it is important to distinguish between secondary and primary adpositions. Although the former are clearly heads in adpositional expressions, it is not so obvious and uncontroversial as has sometimes been assumed that this also holds for primary adpositions, in fact, this issue is the subject of a lively debate. At first sight, this discussion (as in Wunderlich (1991) and Rauh (1996)) appears to differ somewhat from that about articles, auxiliaries, and complementisers in that the main question is not so much whether or not primary adpositions are heads, but whether they are lexical or functional heads. Arguably, however, the concept of a functional head is but one way of formalising the idea that the head characteristics in a given construction are split between a grammatical element, which marks the grammatical function of the construction, and a lexical element, which provides its semantic core. Note also that Zwicky (1993: 306) argues that at least in the case of a subset of primary adpositions (i.e., grammatically used ones), the head features are distributed among the adposition and the noun. In short, there is sufficient evidence to suggest that primary adpositions are not clearly and unequivocally (lexical) heads and that, therefore, constructions involving primary adpositions are not fully compositional with regard to their syntax.

As for semantic transparency, it is generally agreed that primary adpositions such as English in, on, or at are extremely polysemous items and that it is difficult even to account for only a part of their uses, viz., their presumably basic spatial uses, let alone their highly varied extended uses.

In addition to the two criteria just discussed, there are other characteristics which may help diagnosing high entrenchment and lack of compositionality. As mentioned above, high frequency has been linked with high entrenchment. Now, primary adpositions are generally among the most frequent items in their languages. In this regard, it would be instructive to know whether the different behaviour of the eleven primary prepositions in Albanian (of which five interact with definiteness marking on unmodified nouns, while the others don't) correlates with different frequencies. Provided that frequency is indeed linked to entrenchment, the present hypothesis would predict that the most frequent adpositions will be the last ones to succumb to the intrusion of articles into adpositional phrases.

The further exploration of this and related issues is a task for the future. Here the major concern was to establish empirically that the interaction between articles and adpositions is not a sporadic irregularity. It is regular in that it is widespread across languages and in that it occurs only with D-elements and adpositions which are strongly, but not fully, grammaticalised. The crosslinguistic variability of the details of these interactions strongly suggests a historical scenario in which the grammaticalisation of articles begins in core argument positions and only later spreads to constructions with primary adpositions. As a possible answer to the question of why the older, articleless syntax may be retained in adpositional expressions for an extended period of time, it was suggested that constructions with primary adpositions show characteristics of highly entrenched and not fully compositional constructional patterns which isolate them from changes occurring in related, but less entrenched constructional patterns.

Received: 25 October 1997
Revised: 17 June 1998

Notes

An earlier version of this paper was presented at the conference on Functional Approaches to Grammar (Albuquerque 1995) and at the Universität Konstanz. I gratefully acknowledge the helpful comments and suggestions received on these occasions as well as the help and encouragement provided by John Haiman, Paul Hopper, Carl Rubino, Edgar Suter, Angela Terrill, Chikao Yoshimura, and the anonymous referees for Linguistic Typology.
Article use in adpositional phrases

Correspondence address: Institut für Sprachwissenschaft, Universität zu Köln, D-50923 Köln, Germany; e-mail: himmelmann@uni-koeln.de

1. Another syntactic context where articles are commonly omitted are binomial expressions, as in **Herzog moved cast and crew up the river** (Lambrecht 1984: 760).

2. Krámsky (1972), for example, does not discuss this phenomenon although he reports idiosyncrasies of article usage in adpositional phrases for a number of languages, including Portuguese, Spanish, Danish, Swedish, and Rumanian and Albanian (1972: 80, 87, 130–131, 133, 167).

3. For detailed discussion of complex adpositions and their delimitation from spontaneous expressions of similar make-up see, for example, Quirk & Mulholland (1964) and König & Kortmann (1991).

4. Relational nouns (front, behalf, and respect in the above examples) are very often used without an article even if they are clearly nouns (for an exception see Section 3.3). This phenomenon will not be further discussed because complex adpositions are fully lexicalised expressions and our major concern here is with article use in productive adpositional expressions.

5. (ñ) represents a labio-velar nasal. In glisses and figures the following abbreviations are used: abl ablative, acc accusative, act acto, all illative, ben benefactive, caus causative, cl class, com comitative, cont continuous, dat dative, def definite, dir directional, dist distal, do direct object marker, dur durative, erg ergative, exist existential particle, f feminine, gen genitive, hodi hodiernal, impf imperfect, indef indefinite, ins instrument, irr irrealis, lex linker, loc locative, m masculine, n neuter, nom nominative, part participle, part past tense, pl plural, pm predicate marker, pn proper noun, poss possessive, perf perfect, prox proximal, pres present tense, real realis mode, red reduplication, rel relative (particle), rem remote, refl reflexive, sg singular, spec specific, stat static, sup supine, undergoer.

6. For more extensive discussion of relational nouns see Bowden (1992) and Svorou (1994).

7. The following discussion is based on Rauh (1995), making use of her examples. Note that Rauh’s classification is somewhat more detailed, distinguishing five types of prepositions in English.

8. Figure 2 is based primarily on the proposals of Lehmann (1982: 74–106; 1985). See also König & Kortmann (1991), Bowden (1992), and Svorou (1994). It is only concerned with the grammaticisation of relational nouns to adpositions. There are two other major sources for adpositions: adverbs and verbs (cf. Lehmann 1982: 74–106).

9. See Lehmann (1982: 77) for examples and further discussion.

10. Most examples are taken from the *Peer Stories* (Chafe 1980). For further discussion see Himmelmann (1996: 1979: 34–108).

11. In Australian and American Indian languages it is common to find demonstratives specialised in this use. The labels applied to these demonstratives differ widely and include “anaphoric”, “evocative”, and “remember” (cf. Himmelmann 1997: 71).


13. This list is reminiscent of the list of contexts in which specific articles typically do not occur, as discussed by Greenberg (1978: 64–69).

14. The nominal case marking system of Rumanian is extremely reduced. A full paradigm of case forms only exists for personal pronouns. For nouns not marked for definiteness there is only one overt case distinction, namely between nominative/accusative and genitive/dative in the feminine singular paradigm (e.g., *cașă* ‘a house (NOM or ACC)’, vs. *case* ‘a house (GEN or DAT)?). All other nouns without a definiteness marker are unmarked for case. Thus, *lap* ‘a wolf (masculine)’ could be nominative, accusative, genitive, or dative. These nouns are not glossed for case in the following examples. The paradigm of definiteness-marked nouns consists of two overt case forms: one for nominative/accusative (e.g., *lupul* (masculine) or *casa* (feminine)) and one for genitive/ dative (e.g., *lupulai* (masculine) or *casei* (feminine)). Here case is glossed, based on the evidence provided by the personal pronouns.

15. As one of the referees points out, relational nouns in Hungarian complex postpositions are also always marked as definite.

16. Similarly in Romance languages.


18. As a matter of fact, there is one other syntactic environment in which considerable variation occurs with respect to article use, namely non-referential predicate nominals as in She is a teacher (cf. German Sie ist Lehrerin). This environment, however, concerns so-called indefinite articles and is arguably of marginal relevance for the syntax of nominal expressions.

Note that in the case of lexically conditioned variation, the syntactic environment is irrelevant. For example, if the definite article is conventionally used with proper names or the names of (some) rivers and mountains then it is used with these names in all syntactic environments.

19. Furthermore, the Germanic data appear to be more readily amenable to a semantic explanation. This, however, is illusory, as argued in Section 5 below.

20. The marker *ny* is not of direct relevance to the argument in this section and will not be further discussed. It is glossed as definite in the examples. See Naylor (1980) for further discussion.

21. Pawley & Reid (1980: 123, Footnote 8) make exactly the same point, without, however, providing an explicit argument.

22. Tagalog examples for which no source is cited are from the author’s corpus of spontaneous narratives.


24. The paradigm given in Table 6 is only partial in that it excludes plural forms. Furthermore, Barlaan adds a second common noun phrase marker to the paradigm, *na*, which he calls a “definitizer” (1986: 102). However, from the distribution he gives for this element (1986: 99–100) it may be inferred that it is, in fact, a recognitional demonstrative, and hence does not directly bear on the present discussion.

25. That is, there is a statistically significant correlation between the order of V and O and the order of article and noun: in VO languages articles tend to precede nouns while in OV languages they tend to follow.


27. The fact that English demonstratives are inflected for number (*this/these, that/those*) is Cann’s (1993: 45, 51–52, passim) major evidence for his claim that demonstratives require a syntactic analysis different from other determiners, being SPEC of N rather than head of DP (Cann 1993: 60).

28. Even here there is at least one (minor?) difference between article and demonstrative: Some forms of the article fuse with some adpositions (*de -le = du, à -le = au*) but the forms of the demonstrative do not. Furthermore, it has been claimed that ce is well on its way of becoming the new definite article of French while *le* has moved on towards becoming a specific article (cf. Harris 1980 and elsewhere).

29. For further discussion of crosslinguistically widespread differences in the morphosyntactic behaviour of determiners see the flourishing literature on functional
categories in nominal expressions (for example, Szabolcsi 1994, Giusti 1995, or Ritter 1995).

30. But see the remark on the possible role of frequency in this regard at the end of Section 5.

31. In Tagalog, no simple preposition exists for expressing comitative relations (the general purpose preposition sa is not amenable to such use). Instead a variety of periphrastic expressions are used, most of which involve the word (ka)sama ‘company, accompanied, go along with’. Typical examples include the following:

\textit{namansik siyé kasama ng iláng mga tao-ng-bayan}

\textit{REAL.ACT.MOUNTAIN 3SG accompanied GEN FEW.LK.PL PERSON.LK.TOWN}

‘He took the mountains along with a few fellow-townsmen.’ (Bloomfield 1917: 197)

Note that the complement of \textit{kasama} (the fellow-townsmen) is expressed by a genitive construction.

32. Most discussions of such fusions are based on data from the standard (written) language. In colloquial German, much more variation and possibilities exist (like those just given) which are not reflected in the written language. In many instances it is hard to tell whether the fused article is the definite or the indefinite one. Thus, \textit{mit\textit{+}einem} can also be the result of fusing \textit{mit\textit{+}einem} ‘with a’. See Himmelmann (1997: 105–107) for more data and discussion.

33. Plank & Moravcsik (1996: 205) note that SOV basic word order tends to disfavor articles. Inasmuch as SOV correlates with agglutinative case marking, the present observation amounts to the same claim.

34. A variety of explanations may be entertained for this exception, including that of extended contact with German. Note also that the article in Hungarian occurs on the left periphery of the NP while the case markers are located on the right periphery. For an in-depth study of the historical development of the article in Hungarian see Behrens (1989: 159–263). Other apparent exceptions to the universal tendency, such as Mordvin and Basque, are briefly discussed in Himmelmann (1997: 155).

The universal tendency could be strengthened by a more rigorous distinction between (agglutinative) affixes and (phrasal) clitics. With case markers which are phrasal clitics (as in Korean and Japanese), the tendency is actually an implicational universal: if a language has an elaborate paradigm of phrasal case markers, then it will not have articles.

35. For English see, for example, the sections on the peculiarities of article use and development in prepositional phrases in Jespersen (1949: 459–463) and Barber (1976: 226), for French Nyrop (1925: 165–166) and Gamillscheg (1957: 96–98).

36. This position is also found in much other work, including Pawley (1985, 1986) and Hopper (1987). It has also been incorporated into several grammatical frameworks, including Cognitive Grammar (e.g., Langacker 1987) and Construction Grammar (e.g., Goldberg 1995).

37. A third criterion, based on the observation that highly entrenched constructional patterns may provide the model for the creation or reanalysis of related constructional patterns, is omitted from the present discussion.

38. See, for example, Zwicky (1985), Hudson (1987), Hewson (1991), and the contributions in Corbett et al. (eds.) (1993). The position endorsed here—that in the controversial constructions the head features are distributed among different constituents—is, in somewhat simplified form, also the one proposed by Zwicky (1993).

39. For other instances of overdetermination see Plank (1995).

40. Incidentally, the failure to distinguish between primary and secondary adpositions is also one of the reasons why adpositions have somewhat less prominently figured in the headedness debate.

References


