Gradual expansion in the use of the definite article
Checking a theory against the Old Hungarian Corpus

Barbara Egedi & Eszter Simon
Research Institute for Linguistics
Hungarian Academy of Sciences

1 Introduction

The Hungarian Generative Diachronic Syntax Project (2009-2013)
– supported by the Hungarian Scientific Research Fund (OTKA No. 78074)
– located in the Research Institute for Linguistics of the Hungarian Academy of Science

The aim of the project is twofold:
– to digitize all the Old Hungarian records and some selected texts from the Middle Hungarian Period, and to build an on-line searchable historical language corpus
– to reconstruct the syntax of different synchronic systems, and to examine and model the grammaticalization processes

The aim of the present talk:
– to explore and present the possibilities and the limits of checking a linguistic hypothesis against a larger amount of data

Data: five (at least) normalized codices from the Old Hungarian Corpus
The hypothesis to be checked: the expansion in the use of the article

Table 1. Language stages of Hungarian and types of sources

<table>
<thead>
<tr>
<th>Proto-Hungarian</th>
<th>1000 BC – 896 AD</th>
<th>No written documents; reconstructed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Old Hungarian</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Old Hungarian</td>
<td>The age of the Árpád dynasty</td>
<td>Sporadic records, glosses, a few short texts, e.g. <em>Funeral Sermon and Prayer</em> (ca.1195)</td>
</tr>
<tr>
<td>Late Old Hungarian</td>
<td>From around 1370 (time of compilation of Jókai Codex) to 1526</td>
<td>Mainly codices: translations of Latin religious literature + original Hungarian compositions (documents, poems and letters)</td>
</tr>
</tbody>
</table>
| Middle Hungarian  | 1526-1772       | Book printing
 New secular genres appear
 Significant increase in quantity |
| Modern Hungarian | 1772- present day |                                    |
2 The definite article in Old Hungarian

2.1 The data

Modern Hungarian makes extensive use of the definite article

Definite article together with possessives:

<table>
<thead>
<tr>
<th>Modern Hungarian</th>
<th>Old Hungarian</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) a szerzetes-nek a könyv-e</td>
<td>(3) a(z) szerzetes-nek ø könyv-e</td>
</tr>
<tr>
<td>the monk-DAT the book-POSS</td>
<td>the monk-DAT book-POSS</td>
</tr>
<tr>
<td>‘the book of the monk’</td>
<td>‘the book of the monk’</td>
</tr>
</tbody>
</table>

(2) a te könyv-ed  
the your book-POSS.2SG
‘your book’

Definite article together with demonstratives:

<table>
<thead>
<tr>
<th>Modern Hungarian</th>
<th>Old Hungarian</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5) ez/az a könyv</td>
<td>(6) e(z)/a(z) ø könyv</td>
</tr>
<tr>
<td>this/that the book</td>
<td>this/that book</td>
</tr>
<tr>
<td>‘this/that book’</td>
<td>‘this/that book’</td>
</tr>
</tbody>
</table>

About the syntactic structure of the Hungarian noun phrase, see Szabolcsi 1994, É. Kiss 2000

Grammaticalization: distal demonstrative modifier > definite article
⇒ they cannot be distinguished from each other merely on formal grounds: formal and positional equivalence, and functional overlap (e.g. anaphoric use)

(7) a. az kapu  b. az kapu  
the gate that gate

First attestations of the definite article in traditional grammars: pre-articles; transitional word-class; dual nature: “pronoun-article”. (Bakró-Nagy 1999: 7; I. Galassy 1992: 721-722)
For the same problem in Old English, see e.g. Sommerer (2011)

Semantic/pragmatic criteria to identify an article (Himmelmann 2001: 832-834)
– the larger situation use (= first mention of entities that are considered to be unique, hence generally identifiable, in a given speech community)
– the associative-anaphoric use (=first mention of an entity that is not unique per se but with respect to a previously mentioned referent)

For a more detailed account of the issue: Egedi (forthcoming), with further references

2.2 The hypothesis

Search for the regular absence of definite articles in one of the early codices (first half of the Late Old Hungarian period)
⇒ the Gospel of Matthew in the Munich Codex (composed between 1416 and 1441. The survived copy is from 1466)
Aim: an investigation carried out by a manual search on a closed uniform text
– to classify of the articleless noun phrases with a definite interpretation
– to understand why the article is still absent

Theoretical background for the notion of definiteness:

Key concepts:
Identifiability: the speaker signals that the hearer is able to assign a referent for a certain DP.
Definite article is the grammaticalization of the semantic and pragmatic concept of definiteness

Conclusion of the research: The early definite article appears only in the constructions where the referent of the noun phrase is not anchored in another way, thus absent:
– with proper names and with a group of lexemes that describe entities with a prototypically unique referent
– with demonstratives
– possessor expressions
– in case of a generic reading of the noun phrase
– if the noun phrase is an adjunct rather than an argument of the verb

2.3 Absence of article in definite contexts

2.3.1 Proper names

2.3.1 Proper names and nouns with a unique referent

Nouns with special lexical properties → inherently referential
Rigid designators (Kripke 1972)

A group of lexemes (common nouns in present day language use) seem to behave as proper names in Old Hungarian
→ they describe entities with a prototypically unique referent, thus can be identified easily by the hearer/reader, at least in a biblical context, e.g. god, lord, father (referring to God), devil, king, queen, prophet, virgin, heaven, gospel, etc.
→ some of them also show a moderate oscillation between being a proper name or a common noun, and in course of time, they get completely reanalyzed as ordinary common nouns.
→ if modified by an adjective, they tend to have a definite article more frequently than canonical proper names such as personal and place names. (Imre Samu 1953: 357).

(8) és Ø atyá-t senki sem esmerte hanemesak Ø fiú-t
and father-ACC nobody not knew but son-ACC
‘neither knoweth any men the Father, save the Son’ [Matt 11:27]

(9) Mendez kedig lőtt, hogy betelyesednénk, mely mondatott Ø Úrtól
all-this PRT became that be-fulfilled which was-said lord-ABL
Ø próféta miatt, mondván: Íme Ø szűz vall fiat ō méhében, és szül
Prophet through saying behold virgin has son-ACC her womb-INE and brings-forth
‘All this happened to fulfill what had been said by the Lord through the prophet, saying:
Behold, the virgin has a son in her womb, and will give birth…’ [Matt 1:22-23]
(10) Az napot kedig és az időt senki nem tudja, that day-ACC PRT and that time-ACC nobody not knows
sem Ø mennyek angyali, sem Ø Fiú, hanemcsak Ø Atya [Matt 24:36]
neither heaven-DAT angels-PL.POSS nor son but-only father
‘Nobody knows the day and the time, not even the angels of heaven nor the Son, but only
the Father.

2.3.2 Modified by a demonstrative

Demonstratives encode directly accessible reference → necessarily definite
Lyons (1999: 20-21): a kind of ‘matching constraint’, the hearer is instructed to match the
referent of the DP with some object which is either identifiable/visible in the context, or which
is known on the basis of previous discourse

About the syntax and semantics of demonstratives: Alexiadou 2007: 93-130 and Lyons 1999

(11) Az napok2ban jövő Jánus baptista prédikálván Júdeá-Nnak kietlen-é-ben
that days-INE came John Baptiste preaching Judea-DAT desolation-POSS-INE
‘In those days John the Baptist came, preaching in the wilderness of Judea’ [Matt 3:1]
(In Latin: in diebus illis)

(12) mondj, hogy e kövek legyenek kenyerek-ké
Tell that this stones become.they breads-FAC
‘Tell these stones to become bread’ [Matt 4:3]

(13) És im egy kánanea-i némberi az vidék-ből kijövén ivőlt vala
and behold a Canaan-ADJ woman that region-ELA out-coming cries AUX
‘And, behold, a woman of Canaan came out of the same coasts, and cried’ [Matt 15:22]

New pattern: determiner doubling
→ the simultaneous spelling out of the demonstrative modifier and the article
→ appears only in the 16th century and is characteristic of the Middle Hungarian Period

(5’) az a könyv
that the book
‘that book’

2.3.3 Possessive structures

Possessive constructions → prototypically definite; the referent of the possessed noun is
existentially presupposed and identified via its relation to the referent of the possessor
(Koptjevskaja-Tamm 2001: 964)

No determiner appears on the head noun!
Pronominal possessors

(14) És elhozaték egy tálnyér-on Ø ō fej-e, és and was.brought a platter-SUP his head-POSS.3SG and az lány-nak adaték: és vivé Ø ō any-já-nak the girl-DAT was.given and carried her mother-POSS.3SG-DAT

‘And his head was brought on a platter and given to the girl, and she carried it to her mother’ [Matt 14:11]

Nominal possessors

(15) az gyermek-nek lelk-é-t the child-DAT soul-POSS-ACC ‘the soul of the child’ [Matt 2,20]

(16) az papok fedelm-i-hez the priests chief-POSS.PL-ALL ‘to the chiefs of priests’ [Matt 26:57]

The determiners appearing in phrase initial position always belong to the possessor noun phrase only, and, accordingly, are subject to the same distributional rules as in non-possessed contexts.

(17) Ø Isten-nak igé-jé-t [Matt 13:20]

(18) az ország-nak igé-jé-t [Matt 13:19]

God-DAT word-POSS-ACC

the kingdom-DAT word-POSS-ACC

‘the word of God’

‘the word of the kingdom’

(19) Ø ország-á-nak evangéliom-á-t [Matt 4:23]

kingdom-POSS-PL-NOM gospel-POSS-ACC

‘the gospel of his kingdom’

(20) az ország-nak evangéliom-a [Matt 24:14]

the kingdom-DAT gospel-POSS-NOM

‘the gospel of the kingdom’

2.3.4 Generics

Reference is made to a kind rather than an individual

Generic noun phrase: reference is made to the entire class of entities of which the denotatum of the noun is a member. (Alexiadou 2007: 175, for general discussion: Carlson and Pelletier 1995)

Generics ~ proper names

Generic reading: without article ↔ individual reading: with article

(21) Tahát felkelvén parancsola az szelek-nek és az tengér-nek, so up.getting commanded the winds-DAT and the sea-DAT és lőtt vala nagy csendesség. Bizony az emberek and became AUX big silence verily the men csudálkodnak vala, mondván: Minemő ez, mert Ø szelek are.amazed AUX saying what.kind this that winds és Ø tengér engednek neki? and sea obey.they to.him

‘So he got up and commanded the winds and the sea, and it turned very calm. The men were amazed, saying: “What kind (of man) is this, that the winds and the sea obey him!”’ [Matt 8:26-27]
(22) Látván kedin áz gyölekezetek, félemének és dicsővejték Istent, seeing the assembled awed and praised God
ki adott ilyen hatalmat  O emberek-nek. [Matt 9:8]
who gave such power men
‘When the crowd saw this, they were awed and praised God, who had given such authority to men’

(23) Elég  Ø tanejtvány-nak, hogy legyen, miként Ő mestere,
    enough student that be like his master
és  Ø szolgá-nak, hogy legyen, miként Ő ura. [Matt 10:25]
    and servant that be like his lord
‘It is enough for the student to be like his teacher, and the servant like his lord’

2.3.5 Non-arguments

Argumenthood: syntactic reflex of the concept of referentiality

In Modern Hungarian: Non-referential nominal expressions → in the verb-modifier position or in structural focus position
In Old Hungarian: Non-arguments and focus constituents are optionally in post-verbal position; non-argument adjuncts can remain unspecified

3. The definite article in Old Hungarian – spreading

In the Middle Hungarian Period (from the 16th century onward) the definite article appears in new contexts:
– co-occurring with demonstratives
– preceding a possessed noun with dative-marked possessor

Manual checking of the NT loci in a later Old Hungarian ms. Codex Jordánszky (1516-1519) → expansion in article use with generic NPs and before possessive pronouns

Aims of Corpus checking:
– to demonstrate the proportional increase in the use of the article within the Old Hungarian Period
– to find out in which context(s) it took place earlier
Table 2. The codices used for checking

<table>
<thead>
<tr>
<th>Codex</th>
<th>Date</th>
<th>Content</th>
<th>Level of text-processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jókai Codex</td>
<td>a.1370/c.1448</td>
<td>life and deeds of Saint Francis of Assisi</td>
<td>morpho-syntactically annotated</td>
</tr>
<tr>
<td>Vienna Codex</td>
<td>a.1416/c.1450</td>
<td>books from the Old Testament</td>
<td>normalized</td>
</tr>
<tr>
<td>Guary Codex</td>
<td>before 1495</td>
<td>texts about righteous life, virtues, etc. written for nuns</td>
<td>morpho-syntactically annotated</td>
</tr>
<tr>
<td>Booklet on the Dignity of the Apostles</td>
<td>1521</td>
<td>A tractate about the apostles (with many dialogues)</td>
<td>morpho-syntactically annotated</td>
</tr>
<tr>
<td>Kazinczy Codex</td>
<td>1526-1541</td>
<td>sermons, legends, parables</td>
<td>normalized</td>
</tr>
</tbody>
</table>

Table 3. The proportion of definite articles in five Old Hungarian codices

<table>
<thead>
<tr>
<th>Codex</th>
<th>a/az</th>
<th>Tokens</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jókai Codex</td>
<td>550</td>
<td>22733</td>
<td>2,42</td>
</tr>
<tr>
<td>Vienna Codex</td>
<td>2225</td>
<td>54423</td>
<td>4,09</td>
</tr>
<tr>
<td>Guary Codex</td>
<td>1387</td>
<td>20288</td>
<td>6,84</td>
</tr>
<tr>
<td>Booklet</td>
<td>621</td>
<td>8745</td>
<td>7,10</td>
</tr>
<tr>
<td>Kazinczy Codex</td>
<td>2042</td>
<td>23591</td>
<td>8,66</td>
</tr>
<tr>
<td>SzIT (Modern)</td>
<td>68562</td>
<td>616131</td>
<td>11,12</td>
</tr>
</tbody>
</table>

NB. The hits always include the distal demonstratives, but it does not effect the proportional increase.

Spreading in various contexts

Table 4. Increase in the use of the article in specific contexts

<table>
<thead>
<tr>
<th>Context</th>
<th>Method</th>
<th>Increase in article use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns with a unique referent</td>
<td>One by one checking of lexical items in the Corpus</td>
<td>slightly; inconsistent results</td>
</tr>
<tr>
<td>Demonstratives</td>
<td>Automatic query</td>
<td>none</td>
</tr>
<tr>
<td>Possessives (pronominal)</td>
<td>Automatic query</td>
<td>√</td>
</tr>
<tr>
<td>Possessives (nominal)</td>
<td>Automatic query</td>
<td>none/minimal</td>
</tr>
<tr>
<td>Generics</td>
<td>Not possible automatically</td>
<td>???</td>
</tr>
</tbody>
</table>
Table 5. Examples for nouns with a unique referent:

<table>
<thead>
<tr>
<th>Codex</th>
<th>Jókai</th>
<th>Vienna</th>
<th>Guary</th>
<th>Booklet</th>
<th>Kazinczy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article use</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>hold ‘moon’</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>föld ‘earth’</td>
<td>24</td>
<td>11</td>
<td>143</td>
<td>47</td>
<td>14</td>
</tr>
<tr>
<td>pokol ‘hell’</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>isten ‘god’</td>
<td>176</td>
<td>2</td>
<td>264</td>
<td>2</td>
<td>141</td>
</tr>
</tbody>
</table>

Barbara Egedi: egedib@yahoo.com
Eszter Simon: simon.eszter@nytud.mta.hu
Corpus query tool: http://ohc.nytud.hu
Downloadable publications: http://www.nytud.hu/oszt/elmnyelv/mgtsz.html

References

