Frančiška Trobevšek Drobnak   UDK 811.111'04'367.625:821.111.09Chaucer G.
University of Ljubljana[*]

DISTRIBUTION OF INFINITIVE MARKERS
IN CHAUCER'S *CANTERBURY TALES*

INTRODUCTION

1. Middle English infinitive markers: origin and distribution

1.1 The six marking patterns of the Middle English infinitive, [-to, -(e)n], [-to, +(e)n], [+to, -(e)n], [+ to, +(e)n], [+for to, -(e)n], and [+for to, +(e)n], are combinations of three formal Middle English infinitive markers: the suffix –**e(n)**, the particle **to** and the compound marker **for to**.

The Middle English suffix **–(e)n** arose from the combination of *-**ono**-, the Indo-European affix of the verbal noun (*nomen actionis*), and ***-m,** the nominative/accusative case ending (Kisbye 1971: 1). The suffix survived as the ending **-an** of the Old English uninflected infinitive, in some verbs, such as *beon, seon, gan, don, fon*, reduced to **–n**. In Northumbrian, the final **n** of the infinitive ending disappeared already in Old English, in Midland dialects by the year 1300, in the south it survived until the 15th century (Wright 1928:72). The loss of the word-final **n** in unaccented position affected other grammatical forms besides the infinitive (e. g. the present plural indicative and subjunctive forms of verbs, the plural of weak nouns and adjectives), but not necessarily at the same time. In Wycliffe's Bible translation (1378), the final **n** consonant is lost in disyllabic infinitives, but still preserved in the plural present indicative forms and in the past participles of strong verbs. After the loss of **n**, the subsequently word-final **e** [ə] ceased being pronounced, first in Scottish and northern dialects (by the middle of the 13th century), then in Midland dialects (by the middle of the 14th century), and, latest of all, in southern dialects, especially in Kent. The exact dates are difficult to determine, since **e** was usually retained in writing. In Chaucer's poetry, the word final **e** was generally pronounced in disyllabic words with a long stem-syllable at the end of the line, and mostly silent in other positions. It is safe to assume that by the end of the 14th century the word-final **e** had been lost in all forms and in all dialects (Wright 1928, ibid.).

The Middle English infinitive marker **to** arose from the directional preposition/adverb meaning 'towards', which together with the dative case of the verbal noun in **–enne** functioned as the inflected infinitive (supine) in Old English. The variant suffix **–anne**, which appeared probably through analogy with the suffix **-an** of the

---

[*] *Author's address*: Filozofska fakulteta, Oddelek za anglistiko in amerikanistiko, Aškerčeva 2, 1000 Ljubljana, Slovenia. Email: franciska.trobevsek@ff.uni-lj.si

uninflected infinitive (Kisbye 1971: 7), coalesced with the suffix **-an** of the uninflected infinitive at an early stage, especially in northern dialects. In Northumbrian poetry, **to** was followed by the bare form of the infinitive from the earliest days. The inflected infinitive was of relatively limited occurrence in verse and quite rare in prose. It was used with deontic verbs *agan* and *habban,* and frequently in complementation of adjectives *gearu* 'ready', *geornful* 'eager' and *eaðe* 'easy'. It was very rare in nominal functions, except as the subject of an impersonal verb. The inflected infinitive lost its supine function, and **to** was "reduced to a meaningless infinitive sign" (Kysbie 1971: 2) already in Old English. According to some linguists, however, **to** is not part the infinitive, but rather a separate syntactic word, a subordinator introducing the infinitival non-finite clause (Huddleston and Pullum 2002: 84).[1]

The **for to** marker appeared already in Old English. Initially, it served as an indicator of purpose, replacing the old inflected infinitive in the supine function, but soon followed the course of its predecessor, and became an ordinary infinitive marker. Around the year 1300, the use of the **for to** infinitive reached its peak, then it declined and survived only in some northern regional dialects (Hughes and Trudgill 1966: 116). Kaartinen and Mustanoja believed that the choice between the **to** and the **for to** marker of the infinitive was "largely dictated by metrical conditions" (1958: 179). Chaucer used both markers, with some verbs, like *beginnen, desiren, hopen, lernen*, even all three infinitive markers without any functional/semantic distinction (Fisher 1992: 172).

The present-day distribution of the bare infinitive and of the to-infinitive must have been reached quite some time before 1500. The comparison of the use of the two forms in the Old English Gospel according Saint Mark (Corpus MS, 10th century), in Wycliffe's Bible translation (1378), in King James Bible (1611) and in *The New International Version of The Holy Bible* (1982) reveals that the distribution of these two forms in Wycliffe's text is the same as in the two modern English translations[2] (Trobevšek 2003: 104).

1.2 Most discussions on the distribution of different infinitive markers in Middle English focus on the proliferation of the particle **to** and its transformation from a supine to a common infinitive marker.

Callaway (1913) investigated the correlation between the semantic type of the matrix verb and the type of complementation in Old English. He divided all verbs into three groups: those that occur only with the bare infinitive, those that occur only with the to-infinitive, and those that occur with either of the two forms. He dis-

---

[1] That **to** is as a separate syntactic word and not part of the infinitive is, according to Huddleston, evident from the fact that it can stand on its own in elliptical constructions (*I haven't read it yet but I hope to shortly*), need not be repeated in coordination (*I want to go out and get some exercise*) and can be separated from the verb by an adverb (*I'm trying to gradually improve my game*).

[2] No *for to* infinitives were found in the analysed Wycliffe's text.

covered that all semantic groups of verbs, with the exception of modal verbs and verbs of perception, appear on all three lists, and assumed that the variation must be accounted for on syntactic grounds. Comparing the nominal and the infinitival complementation of verbs, Callaway concluded that verbs complemented by accusative objects are more likely to be followed by bare infinitives, and that verbs complemented by objects in the dative or genitive case are more likely to occur with to-infinitives. Verbs that can be followed by either form are verbs that can be followed by objects in different cases (Callaway 1913: 63).

Kaartinen & Mustanoja (1958) analysed late Middle English texts and discovered correlation between the form of the infinitive and the intimacy of the relationship between the matrix verb and the infinitive, including the physical distance between them. Quirk & Svartvik (1970) deduced the same from the studies of Chaucer. The degree of intimacy, as understood by these and other authors (Sanders 1915, Ohlander 1941), is proportional to the degree of grammaticalization of the matrix verb. It is highest when the matrix verb is void of referential meaning, as in the case of modal and other auxiliaries. The same structurally based approach is advocated by Warner (1982: 116ff), who ascribes the propensity of modal verbs to bind with bare infinitives to their auxiliary status. According to Plank (1984: 339), the same tendency is at work when contracted verbal forms *wanna, gonna, bounta, gotta* govern bare infinitives.

Fisher (1992) argues that (for) to-infinitives were preferred when the activity was perceived as taking place sometime in the future. It is by this future orientation, reminiscent of the original supine value, that the to-infinitive stands in contrast to the present participle in the complementation of verbs such as *remember, stop, try* etc. to this day (Biber & al. 1999: 693–739).

Los (1998: 1–36) departs from the common assumption that the to-infinitive replaced the bare infinitive in Middle English. Instead, he suggests that it expanded as an alternative to subjunctive that-clauses, especially those expressing intention, purpose or volition. The ratio of to-infinitives to that-clauses in the prose texts of the Helsinki Corpus stayed the same throughout the Old English period, but changed dramatically from 23% to 74% in the transitional period from Old English to Middle English. According to Los, this change is far more drastic than the change of the ratio of to-infinitives to bare infinitives. The probability that the decrease in that-clauses is unrelated to the increase of to-infinitives is extremely low (Los 1998: 28). The comparison of the Gospel according Saint Mark in the Corpus MS and in Wycliffe's Bible reveals that 17,8% of infinitives in Wycliffe correspond to subordinate that-clauses in the Corpus MS, while no Old English infinitive structure is rendered as a that-clause in Wycliffe (Trobevšek 2003: 104).

1.3 In contrast to the many studies on the distribution of (for) to- and bare infinitives in Middle English, not much attention has been paid to the marker -(e)n. Its decline is described in terms of the general loss of –n and, subsequently, of -e in final unaccented position (see 1.1). With the exception of the commonly accepted view that "when the next word began with a vowel, the –n was run on to it" (Wright 1928: 112),

the factors enhancing or hindering its disappearance have not been mapped out. In the absence of functional motivation,[3] these can be explored in the light of the theories of linguistic *iconicity*, *frequency asymmetries* and *naturalness*.

2. LINGUISTIC ICONICITY, FREQUENCY ASYMMETRIES AND NATURALNESS

2.1 The notion of linguistic iconicity has been introduced by cognitive linguists as "the intuition [...] that the structure of language reflects in some way the structure of experience" (Croft 2003: 102). The most frequently invoked subtypes are *iconicity of quantity, iconicity of complexity* and *iconicity of cohesion*, as defined in the following maxims:

(1) Greater quantities in meaning are expressed by greater quantities of form.
(2) More complex meanings are expressed by more complex forms.
(3) Meanings that belong together more closely semantically are expressed by more cohesive forms. (Haspelmath 2008b: 185–186).

Iconicity of complexity has been defined as correlation between "marked meanings" and "marked forms" (Jakobson 1966: 270), between the "semantic complexity" of a sign and its "phonological representation" (Lehmann 1974: 111), between "a larger chunk of information" and a "larger chunk of code" (Givón 1991), between "morphological expression" and "conceptual intensity" (Haiman 2000: 283). While semantic complexity (markedness) is measured mostly by the number of features needed to describe the meaning of an expression (Lehmann 1974), formal markedness generally means "overtly or more transparently expressed". Typical examples of isomorphism between semantically (more) complex and formally (more) marked values of grammatical categories are:

PLURAL number > SINGULAR number (*girl-s > girl*)
OBLIQUE cases > NOMINATIVE case (*children's > children*),
FEMININE gender > MASCULINE gender (*lion-ess > : lion*)
PAST tense > PRESENT tense (*work-ed > work*)
NEGATIVE propositional polarity > AFFIRMATIVE polarity (*isn't > is*)
IMPERFECTIVE aspect > PERFECTIVE aspect (*was writing > write*) etc.

According to Givón (1985: 189), iconic structures are easier to process than non-iconic structures: "All other things being equal, a coded experience is easier to store, retrieve, and communicate if the code is maximally isomorphic to the experience".

---

[3] In this paper the term *functional/semantic motivation* refers to the choice of a particular form which is based on some lexical or grammatical meaning, e. g. the choice of *(for)to-infinitive* in the supine function, or with (consistent) future orientation.

2.2 Most phenomena to which iconicity of quantity, complexity and cohesion have been applied can be explained (also) in terms of *frequency asymmetries* and *economy principles*. According to Martin Haspelmath, "the great majority of universal morphosyntactic asymmetries are economically motivated. [...] Economical coding is functionally motivated if it occurs with frequently expressed meanings (2008a: 2–3).

Frequent expressions tend to be short in human languages. The correlation, clearly motivated by the speakers' preference for economical structures, is universal. The following order of frequency has been established for different values of grammatical categories across languages (Haspelmath 2002: 238):

NUMBER:	singular > plural > dual
CASE:	nominative > accusative > dative
PERSON:	3rd > non–3rd (1st and 2nd)
DEGREE	positive > comparative > superlative
VOICE	active > passive
MOOD	indicative > subjunctive
POLARITY	affirmative > negative
TENSE	present > future

The main effect of frequency is referential predictability. The relation between the (un)predictability of the referent and the required amount of encoding material is iconic (Givón 1991: 87). The principle of *frequency asymmetry and economy* does not rule out the iconicity of complexity: Greenberg (1966) found out that complex/marked expressions are less frequent universally across languages.

Haspelmath (2008a: 7) claims that frequency-based explanation is "not only sufficient to account for the relevant phenomena, but also necessary, because iconicity of complexity makes some wrong predictions (e. g. FEMALE *widow* and MALE *widow-er*)."[4] Furthermore, it explains why *complementary prototypes*, i. e. "typical associations of a particular value of property/category 1 with a particular value of property/category 2", behave differently than their respective constituents. So, for example, the value [2ndPERSON, IMPERATIVE] constitutes a more frequent association of person and mood than [3rdPERSON, IMPERATIVE], hence a more economical *Run!* than *Let him/us/them run!*. Haspelmath also believes that frequency and economy are among the strongest motivators of diachronic change (2008a: 1).

---

[4] The "wrong" predictions that Haspelmath refers to are "wrong" only if the formal expression is to be isomorphic with semantic and not cognitive complexity. While semantic complexity depends on the number of semantic features embraced in the meaning of the expression, cognitive complexity refers to the ease/difficulty of information processing ("recall" vs "rule application"). Using a mathematical example: the calculus 14 x 13 is more complex than 4 x 13. However, if the computation of 14 x13 is required on a daily basis, the result will become part of the speaker's "recall" (instantly retrievable) list, and consequently processed with less effort than 4 x13, which (still) requires the application of multiplication formulae.

2.3 The notions of markedness, iconicity, frequency and economy of expression are all brought up in the theory of linguistic *naturalness*. It was first formulated as *natural phonology* (Stampe 1979, Donegan 1985) and *natural morphology* (Mayerthaler 1981, Dressler 1987), subsequently extended to syntax (Rydén 1979) and eventually presented as a language-universal theory (Dotter 1990, Dressler 2000, Orešnik 1990). Naturalists currently operate with the terms *naturalness scale*, *sem-values* and *sym-values* (Mayerthaler 1981, 1987; Orešnik 2004). The naturalness scale rests on the assumption that of two morphosyntactic variants one is more natural (<nat) from the speaker's point of view, who is considered the focal point of communication. A typical <nat construction is formally less elaborate, bending to the principle of least effort (Havers 1931: 171). Mayerthaler (1981) divided the naturalness scale into the one referring to the symbolic (formal) properties (sym-values), and the one referring to the semantic properties (sem-values) of linguistic constructions.

On diachronic level**,** the *theory of naturalness* posits correlation of the form and the grammatical environment of a particular expression: formally more elaborate constructions are initially favoured in a relatively complex grammatical environment, and more economical constructions are initially favoured in a relatively simple grammatical environment (Orešnik et al. 1990: 5–11). It is the *inclusion of grammatical environment* that makes the *naturalness theory* distinctive from fundamental Gricean pragmatics (Grice 1975), from iconicity as applied by cognitivists and from frequency-based explanations as proposed by Haspelmath.

3. METHOD

3.1 The analysis of Middle English infinitives in Chaucer's *Canterbury Tales* involved the following steps:

(a) All the infinitive non-finite clauses (INF) in the first 1000 lines of the *Canterbury Tales* were grouped into three samples with two subsets on the basis of their markers:
- Sample 1:   [for to, -(e)n] INF as in *for to beren*
                [for to, -(e)] INF as in *for to bere*
- Sample 2:   [to, (e)n] INF as in *to beren*
                [to, (e)] INF as in *to bere*
- Sample 3:   [(e)n] INF as in *beren*
                [(e)] as in *bere*
(b) In each sample, infinitives and matrix verbs were listed and infinitive tokens were counted.
(c) The ratio of INF containing the marker –**(e)n** and INF without the marker **–(e)n** was computed for each sample. In each subset of samples the number of INF occurring in the middle of the line before a consonant was established (see 1.3).
(d) All INF were defined as to their function in the superordinate clause or phrase. The functions assigned follow the classification in Huddleston and Pullum (2002: 1176) and are:
- Complement in clause structure, as in

To underestimate him would be foolish. [subject - S]

It is not honest to do it. [extraposed subject - ES]

To love is to forgive. [predicative complement - PC]

She wants to leave the country.[catenative complement - CC]

- Complement in phrase structure, as in

I've missed the opportunity to speak out [complement in noun phrase - NC]

She is keen to regain control [complement in adjectival phrase - AdjC]

- Modifier in clause structure, as in

She left at 6 to catch an early train [adjunct of purpose - A]

- Supplement to a clause, as in

He is a charlatan, to put it blunt. [Supp]

(e) The grammatical environment of INF, i. e. the grammatical properties of the superordinate clause, was determined as to:[5]

- the grammatical number of the matrix verb [singular, plural]
- the person of the matrix verb[6] [3rd, non–3rd]
- the tense [present, preterite, future, present perfect]
- the mood [indicative, non-indicative]
- the propositional polarity [affirmative, non-affirmative]
- the type of the matrix verb [stative, non-stative].

(f) The internal structural complexity of INF was determined as to the presence of complements or adjuncts in it.

(g) The ratio (percentage) of individual values of parameters was computed for each subset of samples.

(h) The results were presented cumulatively for all three samples.

(h) The results were discussed from the point of view of predictions ensuing from the principles of iconicity, frequency asymmetries, economy and linguistic naturalness.

4. RESULTS

4.1 SAMPLE 1: INF *with the markers* **[for to, (e)n]** *and* **[for to, (e)]**

Infinitives: **(a)**[7] (for to) seken, tellen, delen, been, helpen, drynken, stonden, werken, seen;

(b) (for to) seke, ryse, festne, yive , have, wynne, saye, be, telle, seche, make, do;

---

[5] The parameters considered are those which are relevant in the determining the congitive complexity of grammatical environment. Generally, these are the marked values of grammatical categories as suggested by the *Prague School* (Jakobson 1932) and by *natural morphologists* (Mayerthaler 1981). For other parameters see also Trobevšek 2009.

[6] The category of person was omitted in SAMPLE 1 on account of the low occurrence of non-third persons (below one percent).

[7] (a) = subsets of infinitives with the ending –n and their matrix verbs; (b) = subsets of infinitives without the ending –(n) and their matrix verbs.

Matrix verbs: (a) longen, wente, is, was (shaply), weren (able), loved, (seemly) was, liketh, was (routhe;
(b) wende, made (forward), hadde (pynn), is, ne was, made, oughte, (myghte) take, is, was nat, (wol) ryde, (wolde) vouche-sauf.

Number of tokens:[8]	[for to, (e)n]	[for to, (e)]
Total	14 (51,9)	13 (48,1)
medial position _C	5 (35,7)	4 (30,7)
Syntactic function:		
Subject	-	1 (7,7)
Extraposed subject	4 (28,6)	-
Predicative complement	-	1 (7,7)
Catenative complement	2 (16,7)	3 (23,1)
Complement in NP	-	4 (30,8)
Complement in AdjP	4 (28,6)	1 (7,7)
Adjunct (modifier)	-	2 (15,4)
Clause supplement	4 (28,6)	1 (7,7)
Grammatical environment		
Plural number	3 (30)	3 (25)
Preterite tense	6 (60)	7 (58,3)
- in stative verbs	6 (60)	3 (25)
Present tense	4 (40)	4 (33,3)
Future tense	-	1(8,3)
Non-indicative mood	-	-
-with modal auxiliary	-	2 (16,7)
Non-affirmative polarity	2 (20)	3 (25)
Stative verb	9 (90)	5 (41,7)
Internal structure of infinitive clause:		
Stative infinitives	4 (28,5)	2 (15,4)
With complement(s)	9 (64,2)	8 (61,5)
With adjunct(s)	2 (14,3)	5 (38,5)

4.2 SAMPLE 2: INF *with the markers* [to, (e)n] *and* [to, (e)]

Infinitives: (a) (to) goon, wenden, ryden, stonden, been, seyn, lyven, sitten, seken, drawen, maken, tellen, doon, speken, seen, yiven, abyden, cursen, talen, swynken;

---

[8] The numbers in brackets represent the ratios (percentages) of the occurrences of individual values in the subsample.

(b) (to) ryse, telle, take, countrefete, poure, yeve, have, make, scol-eye, blame, boyle, rekene, undertake, speeke (2), sende, bigynne, visite, ruste, hyre, wynne, forgyve, drinke, pleye, ryde, reste, springe, kepe, heere, ere, be, do, wreke, dwelle, fighte, devyse;

Matrix verbs: **(a)** longen, lay, began, was, peyned, is, seemed, were, ran, weren, shapen, is, began, hath yiven, han been waitynge, wol nat suffren, sestored;
(b) made forward, peyned, thynketh, studie, was, wiste, lipsed, yaf, were, hadde, nas, nedeth, lefte, is, sette, plesen, suffre, techen, affile, prey, shapen, telle, wente, nere, have, assureth, hath, doon, sente, thoughte, bigan.

Number of tokens:	[to, (e)n]	[to, (e)]
Total	**28 (40)**	**42 (60)**
Medial position, _C	14 (50)	16 (38,1)
Syntactic function:		
Subject	1 (3,6)	2 (4,8)
Extraposed subject	1 (3,3)	1 (2,4)
Predicative complement	1 (3,6)	2 (4,8)
Catenative complement	6 (21,4)	16 (38,1)
Complement in NP	1 (3,6)	4 (9,6)
Complement in AdjP	5 (17,9)	7 (16,7)
Adjunct (modifier)	8 (28,6)	10 (23,8)
Clause supplement	5 (17,9)	-
Grammatical environment		
Plural number	5 (21,7)	4 (9,5)
Non-third person	4 (17,4)	5 (11,9)
Preterite tense	15 (65,2)	26 (61,9)
- in stative verbs	9 (39,1)	14 (33,3)
Present tense	6 (26,1)	13 (31,0)
Non-indicative mood	1 (4,3)	1 (2,4)
Non-affirmative polarity	3 (13,0)	11 (26,2)
Stative verb	14 (60,9)	22 (52,4)
Internal structure of infinitive clause:		
Stative infinitives	10 (35,7)	6 (14,3)
With complements	17 (60,7)	23 (54,8)
With adjunct(s)	11 (39,3)	3 (7,1)

4.3 SAMPLE 3: *INF with the infinitive markers* **[(e)n]** *and* **[(e)]** *– BARE INFINITIVES*

Infinitives:[9] **(a)** swynken, pleyen, been, maken, goon, yeven, stelen, helpen, clepen, techen, doon, tellen, stenten, suffren, synken, fortunen, snybben, tollen, plesen, letten;
(b) ryde, bigynne, sitte, make, daunce, purtreye, write, dresse, falle, carie, kepe, wepe, heere, pace, studie,be, spare, yeve, synge, avaunce, have, selle, hente, preye, herne, teche, endigte, pynche, thynke, assente, knowe, rooste, sethe, broille, frye, swere, laughe, carpe, preche, do, lyve, thresshe, blowe, sowne, take, happe, brynge, clense, byte, speke, crie, fynde, suffre, drede, rede, reherce, spare, seye, stonde, understonde, lye, coste, withseye, paye, shape, areste, drynke, forbere, breke, forbere, ete, speke, go, take, wynne

Matrix verbs: **(a)** koude, myghte, sholde, wolde, kan, bigan, wol, shal, nolde, may, lat;
(b) wolden, wol, koude liste, leet, myghte, sholde, shal, dorste, may, mote, was (levere), gan, nolde, moste, moot, kan, bigan, lat, maketh, herde, bad

Number of tokens:	**[(e)n]**	**[(e)]**
Total	**26 (18,2)**	**117 (81,8)**
Medial position _C	14 (53,8)	26 (22,2)
Syntactic function:		
Catenative complement	26 (100)	117 (100)
- with modal auxiliary	24 (92,3)	108 (92,3)
Grammatical environment:		
Plural number	2 (7,7)	8 (6,8)
Non-third person	6 (23,1)	13 (11,1)
Preterite tense	16 (35,2)	89 (76,1)
Present tense	7 (26,9)	15 (12,8)
Future tense	3 (11,5)	13 (11,1)
Non-affirmative polarity	5 (19,2)	16 (13,7)
Internal structure of infinitive clause:		
Stative infinitives	6 (23,1)	12 (10,3)
With complements	21 (80,8)	45 (38,5)
With adjunct(s)	9 (34,6)	17 (14,5)
Separation from the matrix verb	15 (57,7)	58 (49,6)

---

[9] The second infintives in coordinated constructions in which **(for) to** is ommited, as in *To yeve and lene hym of his owene good...*, are listed in samples 1 or 2.

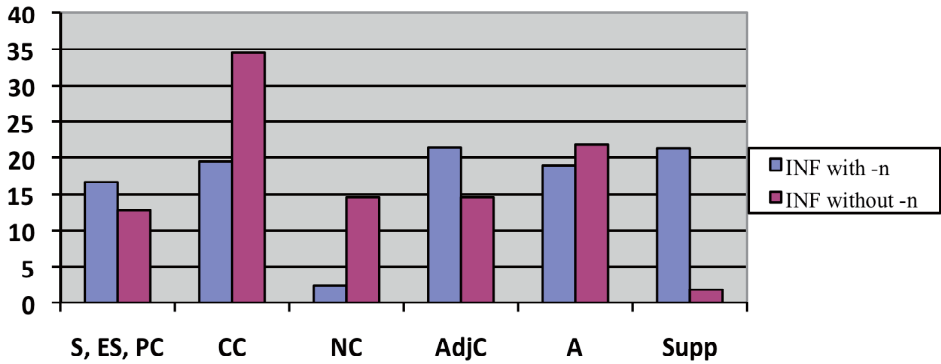## 4.4 Graphic presentation of the results in all three samples

Figure : Syntactic functions[10] of INF with –n and without –n in sample 1 (for to-INF) and sample 2 (to-INF), [11]
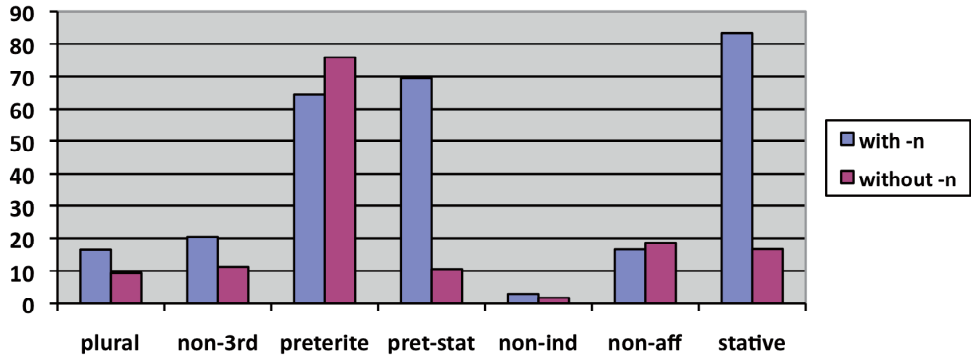
Figure 2: Grammatical environment of INF with –n and without –n in all three samples[12]

---

[10] The results of SAMPLE 3 are not included , since all of the bare infinitives feature as catenative complementa, and more than 90% of them in both subsets are used with modal auxiliaries.

[11] S = subject ; ES = extraposed subject; PC = predicative complement; NC = complement in a noun phrase; AdjC = complement in an adjectival phrase; A = adjunct; supp = supplement to a clause

[12] The results for the non-indicative mood are calculated only for the inflectionally encoded mood and represent only the first two samples, since 92,3 % of infinitives in both subsets of the third sample (bare infinitives) complement modal auxiliaries.
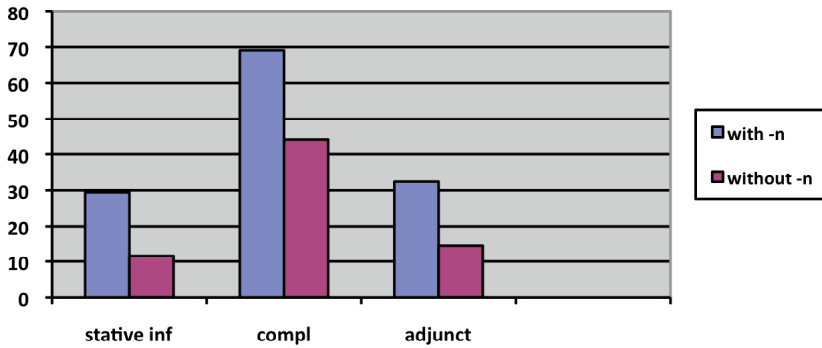
Figure 3: Internal structure of INF with -n and without –n in all three samples

5. DISCUSSION

5.1 The analysis of Chaucer's *Canterbury Tales,* representative of the Middle English language after the 13th century, reveals that the use of different infinitival forms (markers) cannot be explained in terms of their functional (semantic or grammatical) contrastiveness alone. All lexical types of verbs take all six patterns of infinitival markers. As far as matrix verbs are concerned, only modal verbs display a pronounced preference for bare infinitives, but even they show no bias towards the suffix **–(e)n**. The presence/absence of final –(e)n appears arbitrary in all samples. The position before word-initial consonants is neither the only nor the decisive factor of its distribution. As much as 35,7 % of for to-infinitives, 50 % of to-infinitives and 53,8 % of bare infinitives with the final -(e)**n** preserved occur in the middle of the line and before words which begin with a consonant. But arbitrariness does not mean haphazardness. Predictions can be formulated about the (statistical) bias of individual infinitival forms with and without final -(e)**n** for different functions and/or different linguistic environments. Their validation can (a) throw light on the seemingly random distribution of variant forms; (b) test the credibility of the postulates of the theories described in sections 2.1, 2.2 and 2.3.

5.2 Predictions about the relative frequency of infinitive markers ensuing from the theories of linguistic iconicity, frequency asymmetry and naturalness

5.2.1 *Prediction 1: Formally more marked infinitives will be less frequent than formally less marked infinitives.*
There are 240 infinitive clauses in the first 1000 verses of *The Canterbury Tales,* of which

27 tokens (**11,2%**) of **for to-infinitives**
70 tokens (**28,2%**) of **to-infinitives**
117 tokens (**59,6%**) of **bare infinitives**

5.2.2 *Prediction 2: The infinitive marker –(e)n will be more frequent in formally more marked and less frequent infinitive forms.*

The co-occurrence of **(e)n** and other infinitive markers is highest in for to- infinitives and lowest in bare infinitives:

the ratio of –**(e)n** in **for to-infinitives** : 51,9%

the ratio of -**(e)n** in **to-infinitives:** 40%

the ratio of -**(e)n** in **bare infinitives: 18,2%.**

5.2.3 *Prediction 3: The infinitive marker –(e)n will be more frequent in cognitively more marked (complex) and less frequent syntactic functions*

The percentage of tokens of for to- and to-infinitives with the word-final **(e)n** preserved is higher in the functions of **complement in AdjP** and **supplement to the clause**. In the functions of **catenative complement** and **complement in NP**, the proclivity towards dropping the word-final **(e)n** is clearly pronounced both in for to- (23,1% : 16,7%) and to-infinitives (38,1% : 21,4%). The tendency to drop –**(e)n** is strongest in the complementation of modal auxiliaries (81,8% : 18,2%). The results for other syntactic functions are not conclusive on account of the low number of tokens or statistically insignificant difference between respective ratios. The asymmetrical distribution of –**(e)n** confirms, to some extent, the frequency-based predictions, especially those involving *complementary prototypes:* the association of modal auxiliaries and other (catenative) verbs with the infinitive, is more predictable (frequent) than the association of an adjective and a verbal form in its complementation, or a verbal form in a typically nominal function of the subject and predicative complement.

5.2.4 *Prediction 4: The infinitive marker –(e)n will be more frequently preserved in complex grammatical environment.*

The basic postulate the *theory of naturalness*, that more complex grammatical environment favours more elaborate linguistic forms, predicts that infinitives will have the final **(e)n** preserved (to a larger extent) at the following values grammatical categories in the superordinate clause, which are traditionally labelled as more marked (see 2.1 and 2.2):

plural subject
non-third person(s)
preterite tense
non-indicative mood
non-affirmative polarity
stative matrix verb.

The prediction is clearly confirmed in the case of grammatical categories **number, person** and the **[+/-stative]** type of verb. The **preterite** tense of the matrix verb favours, although not consistently in all three samples, infinitives **without** final **(e)n,**

which could be interpreted as (a) inaccuracy of the original assumption that more complex grammatical environment favours formally more elaborate structures, or (b) inaccuracy of the assumption that preterite (past tense) constitutes a more complex (marked) environment than the present tense. The results change significantly, and confirm the original prediction, if the value of a more complex grammatical environment is assigned to the **preterite tense of stative verbs** only.[13] The results confirm the bias of **non-indicative** moods for infinitives with –**(e)n** preserved, but they are statistically inconclusive because of the low occurrence and minimal difference between the two infinitive forms. Contrary to expectations, clauses with **non-affirmative** polarity display no proclivity towards final -**(e)n**.

5.2.5 ***Prediction 5****: The infinitive marker –**(e)n** will be more frequent in structurally more complex INF clauses.*
The presence of the word-final -**(e)n** correlates with the presence of complements in all three samples, which could be interpreted that the presence of overt subject (different from the subject of the matrix verb) and/or objects (complements to transitive verbs) constitutes a complex (marked) grammatical environment. The typical association of intransitive verbs with adjuncts (of place) could also explain why the correlation between the presence of final -(e)**n** and the adjunct in the INF clause is not consistent. To the same extent that final **(e)n** persists longer in the environment of stative matrix verbs, the results reveal a pronounced bias of **stative infinitives** to keep the final -**(e)n.**

5.3 When reviewing the postulates of the theory of markedness, Gricean pragmatics, the principles of iconicity, frequency and economy of expression, and, last but not least, those formulated within the *theory of naturalness*, one fact stands out: that they all provide ample evidence of the same phenomenon, the correlation between the complexity of form and the complexity of content. The same kind of predictions about the asymmetries of linguistic structures can be formulated within each of the theoretical framework discussed. The *theory of naturalness*, however, is the only one that purports to explain the relation between the form of linguistic constructions and their propensity for specific grammatical environment.

Although derived from the analysis of a relatively small corpus, the results prove that, in addition to the correlation between the form and the content of linguistic constructions themselves, the nature of the immediate grammatical environment does affect the choice between formally contrastive but functionally elective linguistic variants. The consistency and conclusiveness of the results depends largely on

---

[13] That preterite tense is typically associated with non-stative verbs, and the present tense with stative verbs, is manifested in the process of creolization. In creole languages the primary tense system depends on the stative or non-stative status of the verb. Bickerton reports that the default tense of the zero form of stative verbs is [-anterior], and the default tense of nonstative verbs is [+ anterior] (Bickerton 1975:461).

the accuracy of the evaluation of what constitutes a higher degree of complexity (markedness) of linguistic constructions and their environment. The traditional scales of markedness can still be applied, but they yield more consistent results if their respective token frequencies and (proto) typical associations are taken into account. The most important contribution of the theory of linguistic *naturalness*, on the other hand, is that it takes the correlation of form and content beyond the boundaries of linguistic constructions themselves, and provides a theoretical framework for the exploration of the correlation of the form of linguistic constructions and their immediate grammatical environment.

References

CHAUCER, Geoffrey/Jill MANN (eds) (2005) *The Canterbury Tales.*New York: Penguin Books.

CHAUCER, Geoffrey (1378–1400) *The Canterbury Tales.* <http://www.librarius.com/cantales.htm>

BIBER, Douglas/Stig JOHANSSON/Geoffrey LEECH/Susan CONRAD/Edward FINEGAN (1999) *Longman Grammar of Spoken and Writeen English.* London: Longmann.

BOOIJ, GEERT E./Chrisian LEHMANN/Joachim MUGDAN (eds) (2000) *Morphologie: ein internationals Handbuch zur Flexion und Wortbildung.* I. Berlin: De Gruyter.

CALLAWAY, Morgan (1913). *The Infinitive in Anglo-Saxon.* Washington: Carnegie Institution.

CROFT, William (^{2}2003) *Typology an universals.* Cambridge: Cambridge University Press.

DONEGAN, Patricia J. (1990) *On the natural phonology of vowels.* New York/London: Garland.

DOTTER, Franz (1990) *Nichtarbitrarität und Ikonizität in der Syntax.* Hamburg: Buske.

DRESSLER, Wolfgang U. (1987) *Leitmotifs in Natural Morphology.* Amsterdam: Benjamins.

DRESSLER, Wolfgang U. (2000) "Naturalness." In: G. Booij et al. (eds), 288–96.

FISCHER, Olga (1992) "Syntax." In: N. Blake (ed.), *The Cambridge history of the English language.* Cambridge: Cambridge Unversity Press, 207–398.

GIVÓN, Talmy (1985) "Iconoicity, isomorphism and non-arbitrary coding in syntax." In: J. Haiman (ed.), *Iconicity in syntax.* Amsterdam: Benjamins.

GIVÓN, Talmy (1991) "Isomorphism in the grammatical code: Cognitive and biological considerations." *Studies in Language* 15/1, 85–114.

GREENBERG, Joseph (1966) *Language universals, with special reference to feature hierarchies.* The Hague: Mouton. (Janua Linguarum, Series Minor, 59).

GRICE, Herbert P. (1975) "Logic and Coversation." In: P. Cole/J. L.Morgan (eds.), *Syntax and Semantics 3: speech acts.* New York: Academic Press, 41–58.

HAIMAN, John (2000) "Iconicity." In: G. Booij et al. (eds), 281–282.

HASPELMATH, Martin (2002) *Understanding morphology.* London: Arnold.

HASPELMATH, Martin (2008a) "Frequency vs. iconicity in explaining grammatical asymmetries." *Cognitive Linguistics* 19/1, 1–33.

HASPELMATH, Martin (2008b) "Creating economical morphosyntactic pattern in language change." In J. Good (ed.), *Linguistic universals and language change.* Oxford: Oxford University Press, 185–213.

HAVERS, Wilhelm (1931) *Handbuch der erklärenden Syntax. Ein Versuch zur Erforschung der Bedingungen und Triebkräfte in Sytax und Stilistik.* Heilderberg: Winter

HUDDLESTON, Rodney/Geoffrey K. PULLUM (2002) *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

HUGHES, Arthur/Peter TRUDGILL (³1996) *English accents and dialects.* London: Arnold.

JAKOBSON, Roman (²1966) "Implications of language universals for linguistics." In: J. H. Greenberg (ed.), *Universals of language.* Cambridge, MA: MIT Press, 263–278.

KAARTINEN, Anja/Tauno F. MUSTANOJA (1958) "The use of the infinitive in *A Book of London English 1384–1425.*" *Neuphilologische Mitteilungen* 59, 179–192.

KISBY, Torben (1971) *An Historical outline of the English language.* Aarhus: Akademisk Boghandel

LEHMANN, Christian (1974) "Isomorphismus im sprachlichen Zeichen." In: H. Seler (ed.), *Linguistic worshop II: Arbeiten des Koelner Univfersalienprojekts 1973/4.* Muenchen: Fink, 28–123.

LOS, Bettelou (1998) "The rise of the to-infinitive as verb complement." *English Language and Linguistics* 2/1, 1–36.

MAYERTHALER, Willi (1981) M*orphologische Natuerlichkeit.* Wiesbaden: Athenaion.

MAYERTHALER, Willi (1988) *Naturalness in morphology.* Ann Arbor: Karoma.

OHLANDER, Urban (1941) "A study on the use of the infinitive sign in Middle English." *Studia Neophilologica* 14, 58–66.

OREŠNIK, Janez/Andrej SNEDEC/Karmen TERŽAN/Frančiška TROBEVŠEK DROBNAK (1990) "Introduction to the subsequent three papers in the present volume." *Linguistica* XXX, 5–12.

OREŠNIK, Janez (2004) *Naturalness in (morpho)syntax. English examples.* Ljubljana: Slovenska akademija znanosti in umetnosti.

PLANK, Frans (1984) "The modals story retold." *Stories in Language* 8, 305–364.

QUIRK, Randolph/Jan SVARTVIK (1970) "Types and uses of non-finite clause in Chaucer." *English Studies* 51, 393–411.

RYDÉN, Mats (1979) *An introduction to the historical study of English syntax.* Stockholm: Almqvist & Wiksell International.

SANDERS, Hermann (1915) *Der syntaktische Gebrauch des Infinitivs im Frühmittelenglischen.* Heidelberg: Carl Winter.

SKEAT, Walter W. (ed.) (1871) *The Gospel according to Saint Mark. In Anglo-Saxon and Northumbrian versions synoptically arranges, with collations exhibiting all the readings in all the MSS.* Cambridge: Cambridge University Press.

STAMPE, David (1979*) A Dissertation on Natural Phonology.* New York/London: Garland.

STAMPE, David (1982) *The Holy Bible. New International Version.* London: Hodder & Stoughton.

STAMPE, David (1811) *The Holy Bible. Containing the Old and New Testaments set forth in 1611 and commonly known as the King James Version.* New York: American Bible Society, 1811.

TROBEVŠEK DROBNAK, Frančiška (2003) "Linguistic variation and change: Middle English infinitive." *Acta Neophilologica* 36/1–2, 103–114.

TROBEVŠEK DROBNAK, Frančiška (2009) "The interaction of form and content: syntactic constructions and grammatical environment." *Studia Anglica Posnaniensia* 45/2, 145–159.

WARNER, Anthony R. (1982) *Complementation in Middle English and the Methodology of Historical Linguistics.* London: Croom Helm.

WRIGHT, Joseph/Elizabeth M. WRIGHT (²1928) *An Elementary Middle English Grammar.* Oxford: Clarendon.

WYCLIFFE, John (1378) *Bible Translation.* <http://wesley.nnu.edu/biblical_studies/wycliffe>

Abstract

DISTRIBUTION OF INFINITIVE MARKERS IN CHAUCER'S *CANTERBURY TALES*

Three formal markers occur in Middle English infinitives: the suffix *–e(n)* from the Old English *–an* of the uninflected infinitive, the particle *to* of the Old English inflected infinitive, and, as a new Middle English formation, the *for to* marker. The six marking patterns of the infinitive were: [+to, -(e)n], [-to, +(e)n], [+to, -(e)n], [+ to, +(e)n], [+for to, -(e)n], and [+for to, +(e)n]. After the 13th century, they displayed no consistent semantic or grammatical contrast. The semantic content of verbs, the type of verbal complementation, physical distance from the matrix verb, metric and rhyme patterns have all been considered as potentially instrumental in the actual choice of individual markers. The author of this paper explores to which extent, if at all, the marking of the Middle English infinitives covariates with their respective syntactic functions, as well as with the grammatical properties of their immediate contextual environment. For this purpose, the distribution of infinitive markers, in particular of the marker **–(e)n,** is explored through the analysis of samples taken from Chaucer's *Canterbury Tales*.

Povzetek

DISTRIBUCIJA OZNAČEVALCEV NEDOČNIŠKEGA OBRAZILA
V CHAUCERJEVIH *CANTERBURYJSKIH ZGODBAH*

Ena najbolj opaznih značilnosti srednjeangleškega nedoločnika je, da se pojavlja v šestih prepoznavnih oblikah, ki jih določajo kombinacije treh različnih označevalcev: nedoločniškega obrazila **–(e)n**, ki razmeroma zgodaj prične glasovno šibeti, členice **to**, ki v srednji angleščini izgubi svojo prvotno vlogo namenilniškega označevalca, ter členiške zveze **for to**, ki v srednjeangleški dobi najprej prevzame vlogo namenilniškega označevalca, a razmeroma hitro funkcijsko ošibi in zamre. Vsaj od 13. stoletja dalje raba različnih nedoločniških oblik ni ne slovnično ne pomensko dosledno razmejena. Med dejavniki, ki bi lahko vplivali na izbor ene ali druge nedoločniške oblike, so bile preučevane pomenske in vezljivostne značilnosti nadrednih glagolskih zvez, fizična razdalja med pomožnim ali faznim glagolom in nedoločnikom, metrične lastnosti besedila in podobno. Večina študij se osredinja na razmerje med rabo t. i. golega nedoločnika, ki je označen le z glasovno šibečo pripono **–(e)n**, in nedoločnikom, ki ga označuje bodisi členica **to** ali njena okrepljena dvojnica **for to**, prisotnost ali odsotnost končnega soglasnika v priponi **–(e)n** pa se obravnava (zgolj) kot posledica glasovne spremembe, ki poteka hitreje v položaju pred začetnim soglasnikom naslednje besede. Pričujoči sestavek povzema rezultate analize nedoločniških polstavkov v prvih tisočih verzih *Canterbury Tales*, pesnitve Geoffreya Chaucerja iz konca 14. stoletja. Osnovno vprašanje, ki si ga zastavlja, je, ali in v kolikšni meri se posamezni nedoločniški označevalci, še posebej končni soglasnik v priponi **–(e)n**, sopojavljajo s specifičnim skladenjskimi funkcijami nedoločniškega polstavka v nadrednem stavku ali zvezi, kakor tudi s posameznimi vrednostmi slovničnih kategorij v nadrednem okolju.