

V.L. Spasova

Sofia University St. Kliment Ohridski
Agricultural University – Plovdiv

CORPUS-BASED STATISTICS ON THE OCCURRENCES OF ALTERNATIVE INTERROGATIVE STRUCTURES IN ENGLISH AND BULGARIAN

The article outlines the basic characteristics of English and Bulgarian alternative interrogative structures and presents statistical data on their occurrences in four English and four Bulgarian corpora. Corpora data show that alternative interrogatives are a rare type of structure in both English and Bulgarian. In general, they are more frequent in spoken language than in fiction.

Keywords: subtypes of alternative interrogative structures, occurrences, correlative pairs, or, whether – or, if – or, ili, dali – dali, dali – ili, li – li, li – ili.

I. Introduction

The purposes of this article are to outline the basic characteristics of English and Bulgarian alternative interrogative structures and to provide statistical data on their occurrences in four English and four Bulgarian corpora. The paper gives brief information about the corpora used in this research, the form and subtypes of English and Bulgarian alternative interrogative structures, and the types of coordinators and correlative pairs used to join their constituents.

In the article the term **alternative interrogative structure** is abbreviated to **AIS**.

II. Corpus-based statistics on the occurrences of alternative interrogative structures in English and Bulgarian

1. Corpora used in this research

My study of the occurrences of English and Bulgarian AISs is based on eight corpora – two corpora of English fiction works, two corpora of spoken English, two corpora of Bulgarian fiction works, and two corpora of spoken Bulgarian.

1.1. English Fiction Corpus (EFC, 90 508 word forms) compiled by V. Spasova.

1.2. English Corpus of Fiction Monologue (ECFM, 50 370 word forms) compiled by V. Spasova.

1.3. Charlotte Face-to-Face Corpus of Spoken English (CFCSE, 90 630 word forms). It is part of a larger corpus of spoken data (198 295 word forms in total) included in the Open American National Corpus (Open ANC) [1].

1.4. Switchboard Telephone Corpus of Spoken English (STCSE, 50 476 word forms). It is part of a larger corpus of spoken data (3 019 477 word forms in total) included in the Open American National Corpus (Open ANC) [1].

1.5. Bulgarian Fiction Corpus (BFC, 90 326 word forms) compiled by V. Spasova.

1.6. Bulgarian Corpus of Fiction Monologue (BCFM, 50 508 word forms) collected by Tzvetomira Venkova of Sofia University.

1.7. Corpus of Spoken Bulgarian (CSB–A, 89 959 word forms) collected by Krasimira Aleksova of Sofia University [2].

1.8. Corpus of Spoken Bulgarian (CSB–NV, 50 000 word forms) collected by Cvetanka Nikolova and Tzvetomira Venkova [3].

2. Basic characteristics of English and Bulgarian alternative interrogative structures

2.1. Form, coordinators and subordinators

An English AIS is an interrogative structure whose constituents are joined by means of the coordinator *or* [4, p. 163], or by one of the correlative pairs of subordinators *whether – or* and *if – or* [5, p. 1053].

A Bulgarian AIS is an interrogative structure whose constituents are joined by means of the coordinator *ili* (equivalent to the English *or*), or by one of the correlative pairs *dali – dali*, *dali – ili*, *li – ili*, *li – li*, *dali da – ili da*, *da ... li – ili da*, *da ... li – da ... li* [6, p. 51; 7, p. 122].

As we can see, a larger number of correlative pairs are used in Bulgarian. Unlike their English counterparts which function only as subordinators, the above-mentioned Bulgarian correlative pairs are used both as coordinators and as subordinators [6, p. 342].

The interrogative structure is called **alternative** because it offers a choice between two (or more) usually mutually exclusive alternatives expressed by the constituent units [5, p. 824; 6, p. 51, p. 342; 7, p. 123].

Both English and Bulgarian AISs can be made up either of two (or more) phrasal constituents (ex. 1, 4) or of two (or more) interrogative clauses (ex. 2, 3, 5, 6). In the latter case both clauses can be full (ex. 2, 5), or the first clause can be full and the second elliptical (ex. 3, 6).

In all examples below the coordinators and the correlative pairs are in bold type, while the AISs are underlined with a single line.

(1) *Had this been **a dream, or a day-time vision**?* (ECFM)

(2) *As both Vandervoort and Wainwright knew, there were devices used by criminals to decide **whether a credit card in their possession could be used again, or if it was "hot."*** (EFC)

(3) *And uh, I don't know who thought of it, I don't know if it was me or Jeff or Craig but uh, we uh, we had some cigarettes on us and we were smoking trying to be the big rebels, you know, and we had a lighter. (CFCSE)*

(4) *Kolko bjaha stápalata do párvata ploštadka – sedem ili osem? (BFC)*

'How many steps were there to the first landing – seven or eight?'

(5) *Mi tja e stara moma. Moma li e, razvedena li e, ne znam, no kaza, če e gospožica. (CSB–A)*

'Well, she is a spinster. I don't know if she is a spinster or a divorcée, but she said she was single.'

(6) *Dovečera šte vidim dali Kondov i Berovski sa dejstvuvali pootdelno, ili v sádrúžie. (BFC)*

'We'll find out tonight whether Kondov and Berovski have acted separately or in co-operation.'

2.2. Subtypes of English and Bulgarian alternative interrogative structures

Neither in English nor in Bulgarian linguistic literature are AISs divided into types or subtypes. In my opinion, however, they can be divided into subtypes according to the type of coordinator and correlative pair used to join their constituents. Hence, in English we can talk about three subtypes – *or*-subtype, *whether* – or subtype, and *if* – or subtype of AIS. In Bulgarian we can talk about eight subtypes – *ili*-subtype, *dali* – *dali* subtype, *dali* – *ili* subtype, *li* – *ili* subtype, *li* – *li* subtype, *dali da* – *ili da* subtype, *da ... li* – *ili da* subtype, *da ... li* – *da ... li* subtype.

(7) *Was that (i.e. story) told to you or did you read that? (CFCSE)*

(8) *She debated, driving north, whether to stop off at home on the way to Berkeley or coming back. (ECFM)*

(9) *I read a thing, I don't even remember if it was in the Dallas Site or the Inside one, about, uh, companies allowing you and they said that TI was looking into it to purchase extra vacation days, which I thought sounded like a good idea. (STCSE)*

(10) *Koj, a, Ljubo i Didka ili ti i tatko ti mi se smejat? (CSB–A)*

'Who, Ljubo and Didka or you and your father are kidding me?'

(11) *Ságlasen sám, če Kamenov e bil ubit. No dali tova e stanalo váv vrážka s radiopredavatelja, dali s ubijstvototo na Jakimov, ili po druga pričína? (BFC)*

'I agree that Kamenov was murdered. But did this happen in connection with the radio transmitter, or in connection with Jakimov's murder, or for some other reason?'

(12) *I dali za tova, ili ot sážalenie, no tja započna da plače i da me zaštitava, kato kazvaše, če sám nevinen. (BCFM)*

'And whether because of this or out of pity, but she started crying and defending me saying that I was innocent.'

(13) *Opitvat **li** se (mitničeskite služiteli), **ili** naistina vi iznudvat?* (BFC)
'*Are they (i.e. the customs officers) only trying or are they really blackmailing you?*'

(14) *To se očkva utre **li**, drugiden **li** zastudjavane.* (CSB–NV)
'*A cold spell is expected, I am not sure whether tomorrow or the day after tomorrow.*'

(15) *Čovek ne znae **da** gi sažaljava **li ili da** se smee.* (BFC)
'*You don't know whether to feel sorry for them or to laugh at them.*'

(16) *Da be, az ne znam kakvo da go pravja tova "Bože", **da** go maham **li*** (ot teksta), ***da** stoi **li**?* (CSB–A)
'*Yes, I don't know what to do with this word "God", shall I delete it or shall I keep it the text?*'

Careful examination of the Bulgarian corpora shows that there are AISs which, in my opinion, could be best described as formed by means of *li – k-duma* (i.e. *li – wh-interrogative* word). Although such AISs are not mentioned in Bulgarian grammar books, they display the above-described basic features of AISs and this is the reason why I have decided to treat them as a subtype of AIS on the same footing as all other subtypes. The *li – k-duma* subtype is therefore the ninth subtype of Bulgarian AISs.

In an AIS formed by means of *li – k-duma* the two alternatives are expressed by the constituent to which *li* is attached and by the *wh*-interrogative word. A wide range of *wh*-words can be used to mark the second alternative – *kakvo* (*what*), *kolko* (*how much, how many*), *koga* (*when*), *kakäv* (*what – masc, sg*), *kakva* (*what – fem, sg*). The choice of the *wh*-word depends on the communicative intention of the speaker and the type of information they ask for [8, p. 30]. Yet, there is a strong tendency among native Bulgarian speakers to use *kakvo* (*what*) as a general substitute for the second alternative.

(17) *"Slim" **li** e tova, **kakvo** e?* (CSB–NV)
'*Is this "Slim" or what?*'

(18) *Sega v dvanajse **li**, v **kolko** tam, šte se snemat maskite i šte se vidi koj e (toj), i vsički šte se vidim.* (CSB–NV)

'*Now maybe at twelve or what time was it, we'll take our masks off and we'll find out who he is, and we'll find out who everyone is.*'

(19) *Proletna umora **li**, **kakvo** stava.* (CSB–A)
'*Is it spring fatigue or what's going on.*'

(20) *Mnogo sa golemi nešto nomerata. Sto **li**, petdeseta **li** mjarka beše (toj), **koja**.* (CSB–A)

'*The sizes are very big. Does he take size one hundred, or fifty, or what.*'

3. Statistical data and conclusions about the occurrences of English and Bulgarian alternative interrogative structures in the eight corpora

3.1. Occurrences of English alternative interrogative structures

The corpus-based study of the occurrences of the three subtypes of English AISs leads to the following statistical conclusions:

a) AISs are a rare type of structure in English. The total absolute number of occurrences of all the three subtypes in the four corpora (281 984 word forms in total) is only 239 (i.e. 0.08 %). There are 34 occurrences (0.04 %) in the EFC (90 508 word forms), 11 occurrences (0.02 %) in the ECFM (50 370 word forms), 134 occurrences (0.15 %) in the CFCSE (90 630 word forms), and 60 occurrences (0.12 %) in the STCSE (50 476 word forms).

b) The *or*-subtype is the most frequent of the three subtypes. There is a total of 175 occurrences (0.06 %) in the four corpora. The number of the occurrences of the *or*-subtype is 4.7 times higher than that of the *whether – or* subtype and 6.5 times higher than that of the *if – or* subtype.

c) The *whether – or* subtype is the second most frequent subtype. There is a total of 37 occurrences (0.013 %) in the four corpora. The number of the occurrences of the *whether – or* subtype is 4.7 times lower than that of the *or*-subtype, but 1.4 times higher than that of the *if – or* subtype.

d) The *if – or* subtype is the least frequent subtype. There is a total of 27 occurrences (0.01 %) in the four corpora. The number of the occurrences of the *if – or* subtype is 6.5 times lower than that of the *or*-subtype and 1.4 times lower than that of the *whether – or* subtype.

e) AISs are much more frequent in spoken language than in works of fiction. Out of the total of 239 occurrences, there are 194 in the spoken corpora and only 45 in the fiction corpora. The total number of occurrences of the AISs in the spoken corpora is 4.3 times higher than it is in the fiction corpora.

f) All the three subtypes are more frequent in spoken language than in works of fiction. The number of the occurrences of the *or*-subtype in the spoken corpora (146 occurrences) is 5 times higher than it is in the fiction corpora (29 occurrences). The number of the occurrences of the *whether – or* subtype in the spoken corpora (28 occurrences) is 3.1 times higher than it is in the fiction corpora (9 occurrences). The number of the occurrences of the *if – or* subtype in the spoken corpora (20 occurrences) is 2.9 times higher than it is in the fiction corpora (7 occurrences).

Table 1 shows the total number of occurrences of the three subtypes of English AIS in each of the four corpora used in this research.

Table 1

Occurrences of English alternative interrogative structures in the four corpora

Corpus Subtype of Alternative Interrogative Structure	EFC (90 508 word forms)	ECFM (50 370 word forms)	CFCSE (90 630 word forms)	STCSE (50 476 word forms)	Total (Absolute Number in 281 984 word forms)	Total (%)
<i>or</i>	22	7	106	40	175	0.06
<i>whether – or</i>	5	4	15	13	37	0.013
<i>if – or</i>	7	0	13	7	27	0.01
Total (Absolute Number)	34	11	134	60	239	
Total (%)	0.04	0.02	0.15	0.12	0.08	

3.2. Occurrences of Bulgarian alternative interrogative structures

The corpus-based study of the occurrences of the nine subtypes of Bulgarian AISs leads to the following statistical conclusions:

a) Like in English, in Bulgarian AISs are a rare type of structure. The total absolute number of occurrences of all the nine subtypes in the four corpora (280 793 word forms in total) is only 248 (i.e. 0.09 %). There are 66 occurrences (0.07 %) in the BFC (90 326 word forms), 30 occurrences (0.06 %) in the BCFM (50 508 word forms), 105 occurrences (0.12 %) in the CSB–A (89 959 word forms), and 47 occurrences (0.09 %) in the CSB–NV (50 000 word forms).

b) The *ili*-subtype is the most frequent of the nine subtypes. There is a total of 74 occurrences (0.03 %) in the four corpora. The number of the occurrences of the *ili*-subtype is 1.5 times higher than that of the *li – li* subtype and 1.6 times higher than that of the *li – ili* subtype.

c) Next in frequency come the *li – li* subtype (48 occurrences, 0.02 %), the *li – ili* subtype (47 occurrences, 0.02 %), the *li – k-duma* subtype (39 occurrences, 0.014 %), and the *dali – ili* subtype (30 occurrences, 0.011 %).

d) Very rare are AISs of the *dali – dali* subtype (6 occurrences, 0.002 %) and the *da ... li – ili da* subtype (3 occurrences, 0.001 %).

e) There is only one AIS of the *da ... li – da ... li* subtype. There are no AISs of the *dali da – ili da* subtype testified in the four corpora.

f) In general, AISs are more frequent in spoken language than in works of fiction. Out of the total of 248 occurrences, there are 152 in the spoken corpora and 96 in the fiction corpora. The total number of occurrences of the AISs in the spoken corpora is 1.6 times higher than it is in the fiction corpora.

g) AISs of the *ili* subtype are equally frequent in fiction and in spoken language. The total number of occurrences of the *ili*-subtype in the fiction corpora is the same as that in the spoken corpora – 37.

h) AISs of the *li – ili* subtype are slightly more common in spoken language than they are in works of fiction. The total number of occurrences of the *li – ili* subtype in the spoken corpora (28 occurrences) is 1.5 times higher than it is in the fiction corpora (19 occurrences).

i) The *li – li* subtype is characteristic predominantly of spoken language. There are 40 occurrences in the spoken corpora and only 8 in the fiction corpora. The total number of occurrences of the *li – li* subtype in the spoken corpora is 5 times higher than it is in the fiction corpora.

j) The *li – k-duma* subtype is found only in spoken language. All the 39 occurrences are in the spoken corpora.

k) The *dali – ili* subtype is typical mainly of fiction. Its occurrences in spoken language are only sporadic. The total number of its occurrences in the fiction corpora (26 occurrences) is 6.5 times higher than it is in the spoken corpora (4 occurrences).

Table 2 shows the total number of occurrences of the nine subtypes of Bulgarian AISs in each of the four corpora used in this research.

Table 2

Occurrences of Bulgarian alternative interrogative structures in the four corpora

Corpus Subtype of Alternative Interrogative Structure	BFC (90 326 word forms)	BCFM (50 508 word forms)	CSB-A (89 959 word forms)	CSB-NV (50 000 word forms)	Total (Absolute Number in 280 793 word forms)	Total (%)
<i>ili</i>	25	12	24	13	74	0.03
<i>li – li</i>	4	4	29	11	48	0.02
<i>li – ili</i>	14	5	20	8	47	0.02
<i>li – k-duma</i>	0	0	26	13	39	0.014
<i>dali – ili</i>	18	8	3	1	30	0.011
<i>dali – dali</i>	3	1	2	0	6	0.002
<i>da ... li – ili da</i>	2	0	0	1	3	0.001
<i>da ... li – da ... li</i>	0	0	1	0	1	0
<i>dali da – ili da</i>	0	0	0	0	0	0
Total (Absolute Number)	66	30	105	47	248	
Total (%)	0.07	0.06	0.12	0.09	0.09	

III. Conclusions

The corpus-based study of the occurrences of English and Bulgarian AISs results in the following conclusions:

a) English and Bulgarian AISs have almost the same percentage and number of occurrences.

b) On the whole, AISs are a rare type of structure in both English and Bulgarian.

c) The English *or*-subtype and its Bulgarian counterpart the *ili*-subtype are the most frequent subtypes.

d) In general, in both English and Bulgarian AISs are more typical of spoken language than of fiction.

e) All the three English subtypes are more frequent in spoken language than in works of fiction.

f) In Bulgarian, the *ili*-subtype is equally frequent in works of fiction and in spoken language. The *li – ili* subtype is slightly more common in spoken language than in works of fiction. The *li – li* subtype is found mainly in spoken language, while the *dali – ili* subtype is characteristic mostly of fiction. The *li – k-duma* subtype is found only in spoken language.

g) Although *dali da – ili da* and *da ... li – da ... li* are mentioned in grammar books as correlative pairs joining the constituent units of Bulgarian AISs, corpora data show that AISs of these subtypes are characteristic neither of spoken language nor of works of fiction.

References

1. Open American National Corpus (Open ANC) [Online resource]. – URL: <http://americannationalcorpus.org/OANC/index.html> (accessed 14.03.2013).

2. Aleksova K. Corpus of Spoken Bulgarian. Корпус от разговорен български език. [Online resource]. – URL: <http://folk.uio.no/kjetilrh/bulg/Aleksova/> (accessed 14.03.2013).

3. Nikolova Cv., Venkova Tz. Corpus of Spoken Bulgarian. Корпус от разговорен български език [Online resource]. – URL: <http://folk.uio.no/kjetilrh/bulg/Nikolova/> (accessed 14.03.2013).

4. Huddleston R., Pullum G.K. A Student's Introduction to English Grammar. – Cambridge University Press, 2005.

5. Quirk R., Greenbaum S., Leech G., Svartvik J. A Comprehensive Grammar of the English Language. – Longman Group Limited, 1985.

6. Grammar of Modern Standard Bulgarian. Volume III. Syntax. Граматика на съвременния български книжовен език. – Том III. Синтаксис / БАН. – София, 1983.

7. Ницолова Р. Прагматичен аспект на изречението в българския книжовен език. Народна просвета. – София, 1984. – С. 122–123.

8. Тишева Й. Модели за интерпретация на сложното изречение в българския език. СЕМА РШ. – София, 2000.

В.Л. Спасова

**СТАТИСТИКА РАСПРОСТРАНЕННОСТИ
АЛЬТЕРНАТИВНЫХ ВОПРОСИТЕЛЬНЫХ КОНСТРУКЦИЙ
В АНГЛИЙСКОМ И БОЛГАРСКОМ ЯЗЫКАХ
(на материале КОРПУСА ПИСЬМЕННЫХ ТЕКСТОВ)**

Рассматриваются основные характеристики английских и болгарских альтернативных вопросительных конструкций и приводятся статистические данные по их частотности в четырех английских и четырех болгарских корпусах текстов. Данные корпусов показывают, что альтернативные вопросительные конструкции являются редкими в английском и болгарском языках. В целом они более распространены в разговорной речи, чем в художественной литературе.

Ключевые слова: подтипы альтернативных вопросительных конструкций, распространенность, корреляционные пары, *or, whether – or, if – or, или, дали – дали, дали – или, ли – ли, ли – или.*