

Computer Sampling and Quantitative Analysis in Exploring Secondary Functions of Questions in Speech Genres of Intimate Communication

Liubov Stasiuk^[0000-0003-2893-7201].

Lviv National Polytechnic University, Mytropolyt Andrei str. 5, 79013Lviv, Ukraine

Liubov.S.Stasiuk@lpnu.ua

Abstract. Although the methods of grammatical, lexical, functional and cognitive analysis of its units are widely exploited in linguistic research, in most of the verification and hypothesis testing processes the quantitative and statistical methods are involved. The presented research article examines the role of computer sampling and elements of quantitative analysis in linguistic data sampling and occurrences checks of secondary function questions in speech genres of intimate communication. The merits of the abovementioned methods in the early and crucial stages of the research have been outlined. Among a few advantages to be mentioned are the large scope of sample size and the possibility to draw inferences from relatively small amount of sample about the general totality.

Keywords: Computer Sampling, Quantitative Analysis, Secondary Function Questions.

1 Introduction

The advance of new technologies has had its impact on the way we conduct linguistic analysis of language. The infrastructure has been created to support a large collection and sharing of high-quality on-line language corpora that provide resources for the study of conversational interaction (The Linguistic Data Consortium (LDC)) as well as to conduct a close textual reading of literary texts with computer-based tool sets existing to facilitate the analysis. The text corpus that is electronically stored and processed (e.g. The Gutenberg Project texts) is used by linguists for language analysis and hypothesis testing. The application of computer science, its algorithms to the analysis of large data sets enables the linguist to operate on a large sample size to check occurrences, validate linguistic rules within a specific language or discover the realities of language usage.

Some major developments both in communicative and computational linguistics have brought me to revive the linguistic interest in questions and linguistic forms by which they are expressed in conversational interaction, namely:

Copyright © 2020 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- A) the revival of interest in questions and linguistic forms by which they are expressed in conversational interaction;
- B) the increased interest in conversation and its dynamics spurred partly by advances in speech recognition and natural language processing;
- C) the infrastructure that is capable of supporting a large-scale research, including the development of online corpora;

In order to investigate and identify stable forms of speech, the conditions for their use in the communication process, it is important to choose the appropriate research methodology and methods. The research of questions of secondary function in the speech genres of intimate communication was based on the principles of complex application of general and specific scientific methods and techniques: method of computer sampling, structural method, speech act analysis, pragmatic and discourse analyzes, methods of cognitive-semantic modeling, elements of quantitative and functional analyzes. In addition, general scientific methods have been used, like inductive and deductive methods that contributed to the study in the direction from analysis of specific material to generalizations and conclusions and vice versa; oppositional method in order to differentiate the value of opposed units; the descriptive method that involves methods of inventory, segmentation and classification of language units, a comprehensive presentation of the results obtained and the method of argumentation that helped to logically compare the views of the precursors on the object under study.

The diversity of the aforementioned theoretical, epistemological and methodological approaches employed takes a continuous reformulation and bridge-building across the subfields of linguistics (Syntax, Semantics, Phraseology), branches (Communicative Linguistics, Computer Linguistics, Quantitative Linguistics, Cognitive Linguistics, Discourse Conversational Analyses) and also across the related fields (Psychology, Anthropology, Philosophy and Relationships Studies). The study is conducted in the terms of Communicative Linguistics, when the processes of communication of people using living natural language, as well as all the relevant components of communication are examined. Nonetheless, computer sampling has been the main method used for selecting the actual samples of interrogatives for further linguistic and quantitative linguistic analyses.

It is worth mentioning that breadth and depth of such bridge-building is not reflected in the very limited number of books and journals devoted to research methodologies in Linguistics. In addition, discussions of research methods appear rarely or briefly in the various forms of dissemination of linguistics research, are often missing from linguistics university courses or tend to be only discussed as a part of specific branch of Linguistics.

This research article draws on examples of such bridge-building and opportunities and challenges involved. The main aim of the article is to bring to the fore and make accessible what can be seen as an under-discussed and opaque subject. More particularly, the attempt is made to examine the role of the main methods of selection and verification of data in the study of secondary functions of interrogatives in the speech genres of intimate communication, namely the method of computer sampling and elements of quantitative analysis.

Quantitative language analysis and computational statistics have been involved in question and answer investigations by linguists earlier. In particular, N. I. Holubeva-Monatkina [1] used the matrix method to test the classification hypothesis (the evaluation of 26 criteria as class-forming classification values for questions and answers in dialogic speech). Due to the matrix length it had to be processed by computer (EC-1022, EC-1035) for the purpose of applying several mathematical and statistical methods (the package of applied programs SOMI). The hierarchy of questions, answers and their properties was built based on arithmetic mean calculation and standard deviation was used to test the abovementioned hypothesis. To build the adequate classification, the multiple correlation coefficient was calculated to establish the degree of the connection of each criteria of evaluation of every question and answer with the others in the matrix. Finally, factor analysis and cluster analysis have been employed to achieve simple and rational classification.

Mathematical methods have been used to evaluate the reliability of the obtained results when interrogative sentences pragmatic characteristics in their historic dynamics have been researched. The data collected led to the conclusion about the regular similarities and differences in the development of interrogative sentences in different historical periods (16-20th c.) and I. Shevchenko [2] calculated mean comparison using the formula of the measurement of the square deviation of the mean of the two compared sets.

The presented research article will address the stages of the research of the secondary functions of interrogatives in the speech genres of intimate communication at which the computer sampling and elements of quantitative and statistical methods have been used. The application of methods, data collection procedures, the sampling strategy and interventions performed will be described and the results presented. Finally, the relevance of the methods used to scientific knowledge and future research will be made.

2 Procedure and Results

2.1 Computer Sampling Method

When addressing the problems outlined in my research on secondary functions of questions in speech genres of intimate communication I aimed at conducting the research in line with modern tendencies in Linguistics to base one's research on a corpus of data. It was very important to establish my findings through a systematic collection of data or through empirical research. In other words, a special type of corpus restricted in time, genre and theme was needed. The method of computer sampling allowed me to accomplish that goal and extract the data that was: 1) representative; 2) homogeneous; 3) containing a listing of what speakers of a language actually produce or a kind of performance grammar.

Firstly, the preliminary stage of the computer sampling should be described. The preliminary stage in the sampling process is to clearly define the target data. The data should meet the requirements of the samples in statistical linguistics, which are representativeness and homogeneity.

The method of computer sampling allowed us to satisfy the first criteria of an adequate sample for linguistic analysis. In other words, the application of the method enabled us to collect the corpus of the study, which comprised 17 704 utterances with a question mark. It was done via the computer application created in C++ computer programming language, that automatically selects the utterances with the question mark. These utterances are presented in blocks in minimal context with each block comprising the utterance with the question mark that is preceded and followed by a certain other utterance. The utterances with question marks were extracted from the electronic corpus of fiction texts of English and American writers in Project Gutenberg but not exclusively taken from this source either. The fiction literary texts used to analyze intimate communication are chronologically, thematically and genre restricted to satisfy the requirement of linguistic homogeneity of the sample. The texts are focused on intimate relationships (Maugham S. "Of Human Bondage", Galsworthy J. "The Forsyte Saga", Davidson M. "Fight the Best Fight"). Also, fiction novels related to the literary genres of 'women's literature' and 'romance novels' were analyzed. The central themes of romantic novels are courtship and romantic relationships that develop on the background of a certain mystery (romantic-mysterious novels (Brontë Ch. "Jane Eyre", Fowles J. "French lieutenant's woman") historical events (romantic-historical (Townsend E. "In love and war"), modern customs (modern romantic novels (Napier S. "Price of Passion", Fielding L. "The Bride's Baby"). Somewhat different are themes of the genre of 'women's literature' or 'chick-lit'): the main character's relationship with her friends, family, the combination of career and motherhood in the modern world, problems of social and marital status (Fielding H. "Bridget Jones: the edge of reason", Bushnell C. "Lipstick Jungle").

The example of the results of this sampling are given in the screenshot picture below.

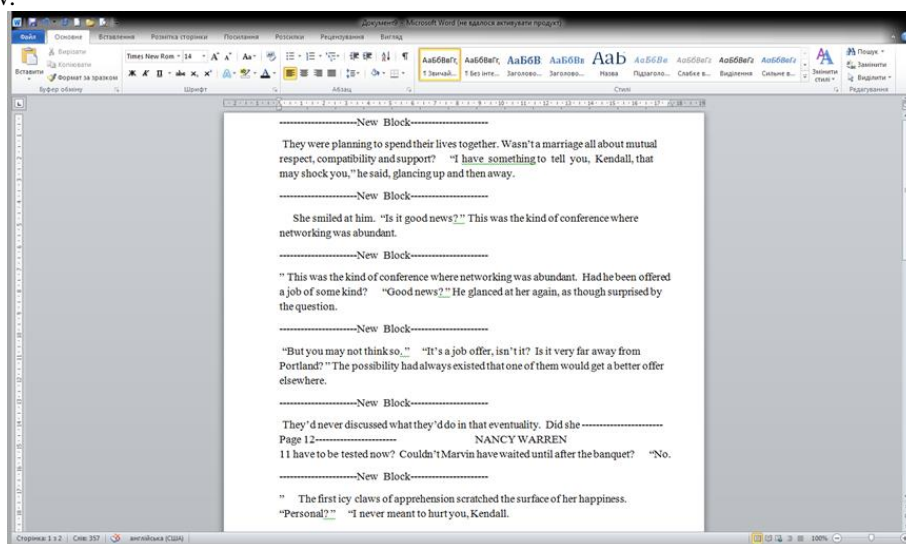


Fig. 1. Blocks of utterances with a question mark obtained as a result of computer sampling

2.2 Quantitative Analysis

At the next stage of the research concerned, 1620 secondary function questions were selected out of the 7269 to be the most representative for different types of linguistic analyses (structural, speech act, conversational, discourse analysis) and quantitative analysis. This selection was made mechanically [3] by selecting from each analyzed text of fiction the even number of questions with a secondary function. Such sampling organization ensures the uniform distribution of the unit under study across the selected set of literary texts.

Some elements of the quantitative data processing method, like frequency counts and determination of percentage correlation, were used to establish the frequency of the analyzed units, above all syntactic-semantic models of secondary function questions, the frequency of secondary function questions in speech acts and speech genres. The quantitative indicators are presented in the form of the following frequency tables: 1. frequency of functioning of structural-semantic patterns of secondary function questions; 2. frequency of functioning of communicative-pragmatic types of questions; 3. frequency of functioning of representative questions; 4. frequency of intrasubjective questions; 5. frequency of use of expressive questions; 6. frequency of functioning of the directive questions; 7. frequency of use of meta-communicative questions; 8. frequency of use of commissive questions; 9. frequency of secondary function questions in intimate communication, 10. frequency of secondary function questions in the speech genres of intimate communication:

Table 1. Frequency of functioning of structural-semantic patterns of secondary function questions

Structure-semantic patterns	Quantity	%
	<i>pattern</i>	
Pattern without interrogative pronouns/adverbs	738	45.6
Patterns with interrogative pronouns/adverbs	469	29
Structurally incomplete patterns	210	13
Patterns with interrogative pronoun and modal verbs	109	6.7
Patterns with tentatives	72	4.4
Patterns with interrogative explicit clichés	7	0.5
Patterns with imputatives	10	0.6
Patterns with infinitive and nexus of deprecation	5	0.2
Total	1620	100

Table 2. Frequency of the functioning of communicative-pragmatic types of questions

Communicative-pragmatic type	Quantity	%
	<i>type</i>	
Representative questions	398	24.6
Intrasubjective questions	382	23.6
Expressive questions	352	21.7
Directive questions	186	11.5
Metacommunicative questions	186	11.5
Commissive questions	72	4.4

Polyillocutionary questions	44	2.7
Total	1620	100

Table 3. Frequency of the functioning of representative questions

Communicative-pragmatic subtype	Quantity	%
	<i>subtype</i>	
Question assertives	144	36.2
Negation questions	94	23.6
Reminder questions	32	8.04
Assumption questions	30	7.5
Appraisal question	26	6.5
Challenge question	24	6.03
Descriptive questions	24	6.03
Confirmation questions	14	3.52
Informative questions	8	2
Predictive questions	2	0.5
Total	398	100

Table 4. Frequency of intrasubjective questions functioning

Communicative-pragmatic subtype	Quantity	%
	<i>subtype</i>	
Descriptive Question	136	35.6
Perplex questions	40	11
Surprise questions	38	10
Censure questions	32	8.4
Anxiety questions	20	5.2
Self-censure question	20	5.2
Negation questions	8	4.2
Projection questions	14	3.7
Self-justification questions	12	3.1
Indignation questions	10	2.6
Justification questions	10	2.6
Self-warning questions	8	2.1
Remorse questions	6	1.6
Self-reassurance questions	4	1.4
Self-brace-up questions	4	1.04
Sympathizing questions	2	0.5
Reminder questions	2	0.5
Regret questions	2	0.5
Total	382	100

Table 5. Frequency of the use of expressive questions

Communicative-pragmatic subtype	Quantity	%
	<i>subtype</i>	
Surprise Questions	88	25
Teasing questions	34	10

.Reproach questions	28	8
Anxiety questions	18	5.1
Indignation questions	18	5.1
Comfort questions	16	4.5
Reprimand questions	16	4.5
Critique questions	14	4
Disappointment questions	13	3.7
Fright questions	10	2.8
Ire questions	8	2.3
Irritation questions	8	2.3
Apology questions	6	1.7
Putdown questions	6	1.7
Blaming questions	5	1.4
Compliment questions	5	1.4
Comical questions	4	1.1
Complaint questions	4	1.1
Catch questions	4	1.1
Pejorative questions	4	1.1
Boast questions	4	1.1
Dread question	4	1.1
Toast questions	4	1.1
Abuse questions	4	1.1
Sympathizing questions	4	1.1
Jovial questions	4	1.1
Expletive questions	4	1.1
Admiration questions	4	1.1
Disbelieve questions	3	1
Contempt questions	3	1
Petulance questions	2	0.6
Jealousy questions	3	1.1
Total	352	100

Table 6. Frequency of functioning of directive questions

Communicative-pragmatic subtype	Quantity	%
	<i>subtype</i>	
Question requests	84	45.2
Proposal questions	42	22.6
Mandative questions	28	15
Invitation questions	16	8.6
Advisive questions	10	5.4
Prescriptive question	2	1.1
Prohibitive questions	2	1.1
Implorative questions	2	1.1
Total	186	100

Table 7. Frequency of the use of meta-communicative questions

Communicative-pragmatic subtype	Quantity	%
	<i>subtype</i>	
Question appeals	48	26
Question politeness formulas	38	20.4
Reflexive questions	34	18.3
Urging questions	22	11.8
Lead-in question	14	7.5
Intrigue question	12	6.4
Control questions	8	4.3
Activizer questions	8	4.3
Greeting questions	2	1.1
Total	186	100

Table 8. Frequency of use of commissive questions

Communicative-pragmatic subtype	Quantity	%
	<i>subtype</i>	
Offer questions	60	83.3
Declinitive questions	8	11.1
Manace questions	2	2.8
Promissive questions	2	2.8
Total	72	100

Table 9. Frequency of the secondary function questions in intimate communication

Communicative-pragmatic type	Quantity	%
	<i>type</i>	
Intrasubjective questions	263	45
Expressive questions	102	17.4
Representative questions	98	16.7
Directive questions	67	11.4
Metacommunicative questions	31	5.3
Polyillocutionary questions	20	3.4
Commissive questions	6	1
Total	587	100

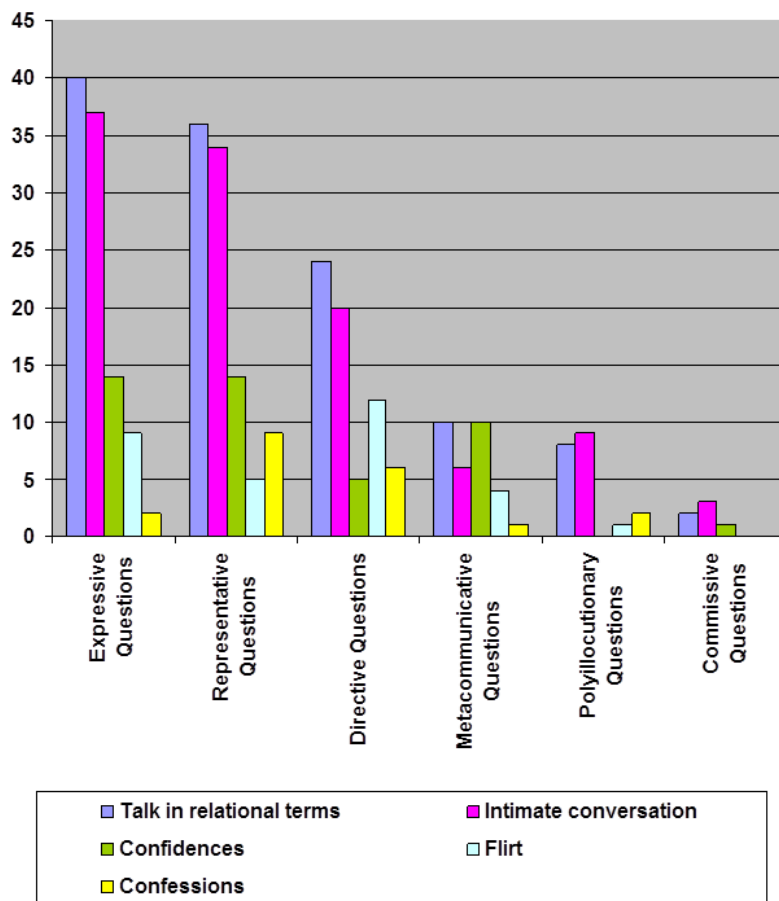


Fig. 2. Frequency of communicative-pragmatic types of questions in speech genres of intimate communication

As we can see from the Table 1, the most frequent patterns in secondary function questions are questions without interrogative pronouns or adverbs, like: Am I my brother's keeper?, Are you insane?, Do you want something to drink?, etc. Second most frequent are patterns with interrogative pronouns, for example: How are you? Where were we? How should I know? Why not? How about a dance? The least widespread are secondary function questions having a structure of infinitive and nexus of deprecation: Me dance? Never!

Speech act and functional analyses allowed us to distinguish communicative-pragmatic types of questions. They are: representative questions, intrasubjective questions, expressive questions, directive questions, commissive questions and meta-communicative questions. It can be seen from numbers in the tables at what speech acts question forms mostly specialize in expressing (table 2) and also some tendencies

in questions to developing new non-interrogative meanings within different speech acts (see tables 3 – 8).The frequency counts point to the fact that secondary function questions are especially productive in expressing representative speech acts or different types of statements and negations and intrasubjective speech acts (see Table 4), which means that we are often thinking in questions. Summing up the information in the tables and a chart, a striking and remarkable conclusion could be drawn about secondary function questions. It is clear that they can express myriads of meanings encompassing all the speech acts and their subtypes. Tables 9 and Fig.2 contain the information on the role of communicative-pragmatic types of secondary function questions in the speech genres of intimate communication. The most productive type of questions with secondary function in these speech genres in interpersonal intimate communication are expressive questions, representative questions and surprisingly directive questions, but only those expressing requests, pieces of advice and promises.

3 Discussion and Conclusions

The minimal context of utterances with question marks, received with the help of computer sampling, is very important for conducting conversational and speech act analyzes when the interactions and sequences of the turns with interrogatives are taken into account. It is impossible to analyze questions to determine their function and communicative-pragmatic meanings without this minimal context, relying on formal properties only.

The pragmatic role of questions and their classification can only be realized through the analysis of textual passages that go beyond sentences. Opposition and comparison methods were used to distinguish between primary and secondary interrogatives (10 435 interrogatives in primary function; 7 269 interrogatives in secondary function). The frequency of questions in the secondary function indicates the prevalence of this property of questions in the corpus selected for analysis. Determining the frequency of manifestation of interrogatives in the secondary function is one of the reasons to consider this property essential for the construction of a dichotomous classification of interrogatives and a detailed analysis of the manifestation of this property.

Consequently, the analysis of lexical-semantic, syntactic, communicative-pragmatic features of secondary function interrogatives, as well as modifiers, descriptors, appeals and personal deixis in these utterances was carried out. The results made it possible to formulate some conclusions regarding the specific functioning of secondary function interrogatives in intimate communication.

The method of observation and contextual analysis were used to study minimal speech situations and to identify the relationship of secondary function interrogatives with contextual features. Thus, the use of computer sampling method helped outline the main contexts in which the use of secondary function interrogatives is triggered, namely: reactive use; irrelevance of the response for the addressee; the addressee knows / assumes that knows the answer, since the answer has only a form of assump-

tion, guesswork; conversion to indirect speech with the help of verbs like tell, ask, demand, postulate, exclaim etc.

A descriptive method and lexico-grammatical analysis were applied, the essence of which is a systematic inventory of linguistic units and an explanation of the peculiarities of their lexico-grammatical structure (See Table 1) and functioning.

At the later stage of the study, a functional method developed within the framework of a communicative-pragmatic approach, which involves the study of language in action, in the process of functioning, was used. The study of communicative-pragmatic peculiarities of secondary function interrogatives in the speech genres of intimate communication has made it expedient to turn to the functional method by which the main communicative-pragmatic types of utterances were distinguished, with the indication of their pragmatic function. The analysis of communicative orientation of questions was made taking into account the illocutionary function of interrogative expression and various parameters of realization of illocutionary force. The relationships between the communicants, their social roles, the time, the environment, in which the communication takes place, the relation of speech act to the interests of the addresser and the addressee and the degree of intensity of the presented illocutionary purpose have been taken into account in the study. In order to elucidate the illocutionary force of the fragments of the studied linguistic material and to typologize illocutionary acts expressed by secondary function interrogatives, speech analysis, induction and deduction techniques were used to create the classification. The quantitative analysis was used to confirm the conclusions about the peculiarities of the functioning and classification of the types of secondary function questions. The results of these counts can be seen in Tables 2 – 8. The frequency tables show what speech act types and subtypes questions specialize in expressing. The high frequency occurrence of question forms with some particular meaning (e.g. expressing surprise) means that they have grown to become specialized forms of expression of this meaning.

In the closing stage, the discourse-analysis method and the elements of conversational analysis have been used to investigate communicative situations of intimate communication in specific language material to reveal their thematic orientation. Intensive deployment of interrogative utterances takes place in compositionally completed speech genres. Intimate communication, in which the intimate speech genres function, is classified as oral, entertaining, informal, interpersonal, intimate-family, intimate-friendly and intimate-romantic, personality-oriented, mixed, initiative-positive and active [5]. Such communication is explored in works of fiction, with all the constituents of the communicative-intentional plan of discourse present. Within the speech genres, questions are speech acts, that is, the addresser transmits to the addressee a performative instruction that influences his or her behavior. The quantitative analysis of frequency occurrences of speech acts of questions in speech genres shows that in 1/3 of the cases of secondary function questions functioning they have appeared to be the part of larger units of discourse than the speech acts, which are speech genres of intimate communication (see Table 9). In less than 1/4 of the cases it was interactional intimate communication (see Table 10).

In conclusion, I would like to encourage linguists to take a wider view of key approaches along the quantitative-qualitative continuum. It is vital to emphasize the

value of mixed methods research and the need to push boundaries of methodologies to incorporate cross-disciplinary perspectives.

The results obtained are relevant to scientific knowledge and future research. Especially important they are to scientists interested in interrogative syntax, language and interaction, computer scientists, human-computer interface designers, to those who deal with human-machine dialogues. The corpus of intimate communication texts has been created based on which a further research into language and intimate interaction could be conducted.

References

1. Holubeva-Monatkina, N. : Voprosy i otvety dialogicheskoy rechi: Klassifikacionnoye isledovaniye. Yeditorial URSS, - M. (2004).
2. Shevchenko I. : Istoricheskaya dinamika pragmaticheskih svoistv angliyskoho VP (16-20 v): Ph.D. dissertation. K., (1999).
3. Hammerl, R., Sambor, J. : O statystycznyh prawach jezykowych. Polskie towarzystwo semiotyczne, Warszawa, (1993).
4. Oakes, M.: Statistics for Linguistics. Edinburg University Press, (2003).
5. Kit, L.: Secondary functions of questions in speech genres of intimate communication in Modern English: Ph.D.dissertation.L., (2014).