Informational Linguistics: Computer, Internet, Artificial Intelligence and Language

Barkovich A.
The Department of Informatics and Apllied Linguistics
Minsk State Linguistic University
Minsk, Belarus
albark@tut.by, 0000-0001-8469-8431

***Abstract*—Modern technological progress is clearly mediated via the informational and communicational conceptualizations, which are first of all of language nature. Interdependence and syncretism of human cognitive activity create unlimited demand for knowledge interpretation of certain semantic format – information. Informational Linguistics is a discipline, dedicated to the interdisciplinary investigation of the specifics of communication contents.**

***Keywords—Informational Linguistics, Artificial Intelligence, computer-mediated communication, Internet, language, semantics, methodology*.**

I. INTRODUCTION

Language functioning in the modern social and technological circumstances requires not only the study of speech samples in separate chronological or spatial frames, but wide systematization of speech practice and its *systematicity* [1]. The current pace of scientific and technological development contributes to significant expansion of speech. This expansion is supported by the *Computer*, revolutionary character of which is compatible with the contrivances like the Wheel or the Fire. Now Linguistics follows the computerization of human interaction. No living language has remained uninvolved in the process of related development of social and cultural sphere: millions of texts are stored on servers operating in the *Internet* [2]. Programs, involved in this process, are enforced with the potential of *Artificial Intelligence*. In this regard, it is necessary to adjust actual methodological framework and correlate the priorities of research practice, including the definition of workable scientific paradigm in order to assure sustainable progress of *human communication* (namely human being should be the only beneficiary of computerization). This *objective* is important in general scientific context as well as in informational aspect of communication.

II. LANGUAGE IN MODERN SCIENTIFIC PARADIGM

The crystallization of *Informational Linguistics* format is predetermined by diffused knowledge status of Natural Sciences – as well as Humanities – in the professional competence of communication researchers [3]. Communication has become computer-mediated, social and cultural development of mankind depends on network technology, but everything seems to remain the same in the world of Humanities as in the Sleepy Kingdom. It looks like *Linguistics* is redundant in the times of technogenic and dynamic progress and is just an anachronism of the scientific romanticism era. Nevertheless there is a paradox: no science is such innovative and useful in the modern world as Linguistics is. With specification: this Linguistics must be "upgradable", or in other words it must be informational.

Probably, today the issues of language system positioning in the symbiosis of knowledge are important as never before. Interdependence and syncretism of human cognitive activity create unlimited demand for knowledge interpretation of some universal format. Apparently, this *format* fits to *Linguistics*. According to E. Benvenist, it is Linguistics that has always been an important source and reserve for scientific development:

> "We observe at the same time that these new methods of linguistics are taken as examples and even as models for other sciences, that the problems of language are today of interest in very diverse and increasingly numerous fields of specialization, and that there is a trend in the research done in the social sciences toward working with the same mind that inspires linguistics" [4].

As practice shows, each individual *computer* as well as *the Internet* and *Artificial Intelligence* cannot exist and develop without *Linguistics*. It is not only the communication saturation of social networks that is relevant here, but their dependence on *language* as the main and the only instrumentality of human interaction as well. Language as a phenomenon is diverse – from gesture to poetry, but one way or another it is a code system, dependent on the means of its "maintenance" – Linguistics. And it is not surprising that all of us are competent in Linguistics: this is indeed the case to some extent.

With that, the Linguistics of modern communication is more *sophisticated* due to increasing "clichédness" and "stereotypy" of speech, equipping it with mechanisms like radio, TV and computer, mediating it by Artificial Intelligence. Principally new kind of meta-descriptions was created – *programming*. With the advent and outspread of a fundamentally new format of language facts functioning – in the communicational networks environment – Linguistics obtained the material of previously unattainable quality and quantity.

With the speech processing instrumentary, such as *Corpus Linguistics* or *Internet search*, the representativeness of language *material* now is often contingent: it is possible to explore either several language units or millions of them in the same way [5]. Step by step

the issues of card files aggregation and their adjustment (computer programs cope with this task successfully and instantly) were replaced by the issues of the formation of research *strategy* and *tactics*. The representative instability of language data suggests its widespread meta-descriptions and *generalizations*. They are extremely needed for entire Linguistics as well as for many neighboring spheres.

Science has always been a *synthetic* phenomenon, which was artificially divided into narrow segments. But in recent years universal methodological techniques have become relevant again, providing new quality of knowledge. Therefore, any ambiguity of current scientific sphere is badly in need of *synthesis* as well as *analysis*: knowledge has to be gathered as well as organized, described and made accessible. Current researches are often carried out on the *boundary* of related disciplines or branches of some discipline. Moreover, interesting and promising findings appear on the border of humanitarian and technical knowledge. Today more often than ever before, the same results of search activity are simultaneously useful for a number of disciplines. Accordingly, the products of research and modeling are often difficult to distribute into certain disciplines and sub-disciplines. It allows identifying their *crossdisciplinary*, *multidisciplinary* or *interdisciplinary* methodological character. Correspondingly research directions gradually merge into a huge common problem domain, what is characteristic for modern science.

Unified methodologies have undeniable perspectives in the context of modern technology and humanity trends, they are already being successfully implemented in such fields as Cybernetics, Web Science, Artificial Intelligence, which are not self-sufficient and homogeneous. Communicational researches are often developed in the trend of synthetic scientific exploration, too. With that, due to their communicational attributability, they are "linguistics-driven," comprehended within the *linguistic* paradigm. On the other hand, contemporary Linguistics is better equipped with various instrumentality and characterized by widened methodological mobility and pliability.

Today language requires consistent and rational *scientific pragmatics* in the systematization and representation of its data. Related kind of Linguistics has to be different: firstly, it should be as correct as possible and only after that – beautiful. The era of computers and network technologies changed the life of mankind and of Linguistics, of course, irrevocably. Long-term perspectivity of any communication tools needs approved scientific support without superfluous limitations. In this way Linguistics is doomed to be *applied science*. It manifests itself in the structuring of such at least syncretic directions as Psycholinguistics, Sociolinguistics, Forensic Linguistics, even Mathematical Linguistics [6] and dozens of others that "appear" annually. However, the novelty of such process is rather superficial: Language, Speech, Communication, and Discourse are the universals that have always existed and have always interfered with the human perception of the world. In such aspect the *informational actualization* of language is an inevitable and consistent stage of knowledge gaining.

So, Linguistics has some exclusivity in Science: it directly or indirectly covers all human activities, including technical and computer fields. With that, all modern technological progress is clearly mediated via the *informational* and *communicational* conceptualizations, which are of language nature, first of all. Many of the technological advances are due to sufficient / insufficient meta-language correctness and the degree of *linguistic competence* of "operators". Namely linguistic or metalingustic support is in strong demand in the innovation processes of science and production.

III. Frames of communicational identity

Today the *essence* of speech phenomena and processes are in focus, as well as their correlation with the world. And the existing views and conceptual representation of such understanding rely namely on *communication*. But there are some "natural" and "artificial" peculiarities. One of them is the *polycode* character of modern speech that is caused by the wide use of multilingual material here. Such are Latin inclusions in English. But English on its own is a much-in-demand source of such inclusions for other codes. Most of the languages are involved and exemplified with related speech practice. Cyrillic languages (such as Russian, Belarusian, Ukrainian) are interesting linguistic subjects in this relation, which allow illustrating speech specificity without the interference of many graphically similar languages (particularly Germanic or Latinic ones), which are to a large extent related. By the way, namely Russian respectively ranks second (after English) in the prevalence on the Internet [7]. Equally indicative is the adaptation of hieroglyphic, Arabic and other authentic written speech to the framework of Latin graphics of browsers and formal languages.

Today, there are principally new speech products – *programs*. They are being reluctantly considered as speech practice, though related activity is based on languages – artificial, formal, but languages. And, moreover, programs are used for processing and meta-description of natural-language speech, causing its *software* dependence. The speech practice due to software support preserves high authenticity and spontaneity that adds special value to the linguistic meta-descriptions. Its main requirement is computerized (or *digital*, *electronic*) format of supported texts, optional – their markup (or annotation, tagging). In their mastering such quality characteristics as accessibility, pre-elaboration (or data pre-processing), adequacy of text processing techniques to the objectives matter, and similar ones are at the forefront.

Additionally, qualitative specificity of this problem domain is complicated by its *meta-textual structuredness* (*hyper-*, *super-*, *inter-*, *sub-* etc.). Current speech activity is global and total: "word" can affect unexpectedly and deeply. Any new meta-description has to take the previous one into account or should affect it, creating *blockchain* quality of communication. A distinctive feature of the relevant problem domain is linguistic capability to involve the widest possible range of illustrative material via computer technology simultaneously. With that, printed texts keep their importance (for example, dictionaries, which are still under-represented in computer ("electronic") format. One way or another, dictionaries are doomed to change the printed format into computerized one and crowd-sourcing.

In the new communicational conditions significant linguistic relevance is provided by strengthening of *above-personal* identity of speech practice, based on *text corpora*, *Internet*, *digital libraries*, *online-dictionaries*.

All such extension of speech practice is more and more separated from its authors. New "personality" of communication is manifested in both organized format of *institutionality* and free format of *"noo-spherity"*, what is especially relevant for *Artificial Intelligence*.

Nevertheless, such instrument as *text corpus* is created for the most of written languages. New instrumentary is already *astronomical* ("hyperscopic") in quantitative measurement. And it needs a new dimension – informational. One more communicational instrument – the Internet, which is "hyper-corpus" – has become a real catalyst for the active development of all current scientific paradigms. It is a nontrivial example of semantic expansion that takes part in the wide replacement of related significance and generalized categories of speech [8]. With that, linguistics of the Internet (or Internet Linguistics) is needed, firstly, as *informationally* determined discipline. Today it is obvious, that any narrowly focused paradigms are not able to catch modern speech dynamics, its manifestations and sense. Now it is the time for new *methodology* of Linguistics, which is based on the statistically verified and semantically (informationally) transparent knowledge [9]. It will unify a wide range of humanitarian competences, it will be supported by computer instrumentality, and it will be complex and multilevel.

Such options of current knowledge mastering can be provided by the universal communicational matter, or "currency" – *information*. The information creates the phenomenological basis for the formation and development of new communicational "religion", or ideology – informational ideology. This ideology, in turn, is indispensible with linguistic paradigm: it is *"linguoinformational"* [3]. The objective side of this aspect strongly requires methodological difinedness: here practice still goes faster than theory. But it is impossible to correct tactics, to reconcile it in a comprehensive manner with a holistic and integrated research strategy without its inner generalized coherence. Generally, linguoinformationality allows preserving the conceptual consistency of the whole communicational activity.

The linguistic maintenance of speech environment gains special identity within the context of dynamic development of contemporary communication. Scientific support of the rapidly expanding communicational content suggests generalizations based on *interpreting*, *representing* and *modeling* of speech. In turn, both active creation of secondary semiotic systems and formalization of natural languages contribute to the involvement of wide linguistic paradigmatics, which allows approved consistent representation of semantics. The universal of *Significance*, mediated in this way, needs thorough theoretical argumentation. But *Significance* as a concept – like Semantics, Syntax and other key terminological cornerstones – has a quite different understanding in Linguistics and Informatics, particularly. Such state of affairs suggests thorough scientific elaboration with unified noncontradictory instrumentarium as well as consentaneous improvement of related categorial structure.

The deepening of the sense scope of speech practice has clear promising opportunities within the considered framework. It is in demand: one of the indicators of such trend is, in particular, the announcement of the creation of *Semantic Web* by T. Berners-Lee and his colleagues [10]. Probably, it is impossible to create it – as well as to create full-fledged Artificial Intelligence – without the representation of *Deep Semantics*. Naturally, solving related problems outside the linguistic paradigm – in the sphere of formal language meta-tools – seems a very attractive goal. However, the experience has proven that all such attempts are still doomed to fail. Meanwhile, the necessary resources are available: this is, first of all, the extension of related linguistic toolkit, for example, with *Meta-Lexical Significance* paradigmatics [11].

Meta-lexical significance is an essential potential of communication, reflecting "deep" semantic identity and functionality according to the complex of own-, sub-, and hyper-lexical significance of language units. A *word*, or more precisely a lexical unit, as a linguistic abstraction, is not a self-sufficient language artifact. It lives on the life force of united *Single Semantic Field of Language*. Meta-lexical categoriality is relevant for communicational modeling: for example, for the creating of a *Conceptual-Dominant Model of Discourse*, *Hyper-Discursive Model* and other models [3]. *Discourse* as a paradigm allows comprehensive investigating of the extra-language specifics of information functioning [12]. Discursive prism provides interdisciplinary value of any speech modeling, making it compatible with all scientific knowledge as well as with its applied specifics. Regardless of the names of the models in M. Minsky, P.N. Johnson-Laird, R.C. Schanck and R.P. Abelson, T.A. van Dijk and others were focused on identifying the content specifics of speech [13; 14; 15; 16].

One way or another, meta-lexical modeling of semantics is in demand in Lexicography, for example, for the representation of contextually dependent variability of semantics. In Machine Translation this fits to extended interpretation of the semantic identity of involved languages. It shows probably most promising results on the material of typologically different languages such as analytical and synthetic ones, for instance. Such elaboration of semantics is perspective, first of all, on the background of *computer-mediated communication*. The functionality of speech in the field of computer-mediated communication has an important benefit: here its identity is supported by powerful meta-language possibilities [17].

Communication as speech environment has necessary dynamism of language functioning, providing its multileveled *"semanticity"*, which is absolutely inherent to information. The linguistic paradigmatics allows real meta-language operating of language semantics, what makes related research practice informationally conditioned. Informational generalizations are really in demand. Probably, it would be advisable to define *information* as an essentially important substance of contemporary communication. Thus, it is advisable to construct any objective description of *creation*, *modification* and *usage* of information on the base of comprehensive and universal methodology. Consistent methodology allows relying on the wide-linguistic argumentation. Therefore, we need the special linguistics of communication – Informational Linguistics.

IV. Linguistic Focus of Information

At first glance there is no difference between *information* and well-known *semantics* – the contents "matter" of speech interaction. But the content in the guise of information has certain aura of newness and authenticity, attracting the imagination of explorers. And

this mysterious appeal is very useful, attracting neophytes. Of course, inexorable logic of the material world suggests that a shiny informational shell of communication is just a cocoon of enchanting butterfly of sense that always escapes from the "entomologists", both amateurs and professionals. But the magic of the art of *word control* has a hypnotic power.

Obviously, Information is one of the key concepts of the communicational sphere: "… hundreds of millions of people engage in information retrieval every day…" [18]. But informational specificity of modern communication is challenging: information is positioned in many linguistic descriptions as a very contradictory substance, particularly in meta-language understanding. In fact meta-language is a secondary semiotic system for any language description in wide sense. But naturally meta-language in the computer-mediated environment is hardly conditioned by the discreteness of formal languages. A meta-language is traditionally based on vocabulary, or lexis. But simultaneously with the assistance of meta-language we encode, transfer and decode some additional significance, some kind of *super significance* – informational significance. It has to be identified otherwise the computer will not use it.

Of course, any semantic relations in computer-mediated knowledge have to be linear and clear. Naturally, after the lexical level Syntax is positioned as the next and the last visible language level. And, accordingly, it is namely in syntax that all the deficit of language significance is looked for. But, initially, syntax cannot substitute all the above-lexical semantics. Really, much of speech semantics depends on the syntagmatics of the software "texts" – programs. But typically right here the misunderstanding of real meta-language functionality begins. Of course, syntax is linguistically sophisticated, moreover, it is hard to simulate it in computing processing. In this way, all the semantic vacuum of meta-descriptions must be associated with something above structural knowledge about language – for instance, information. But, in this way, according to such misunderstanding, information is only declared without any intention to be filled with any useful sense. Of course, it is convenient: "These semantic aspects of communication are irrelevant to the engineering problem" [19]. Naturally, for example, we already do associate many unresolved problems of communication with information today. But information is not just lexical or just syntactical – as well as something else artificially framed in science – it is the essence of all the language involved in communication.

Information as a phenomenon is integrated in communication and is tightly tied with it. And, particularly, computer-mediated information does not match only the traditional understanding of this concept. With that "information" is an internationalism: there were no problems in customizing of the same exponent in most modern languages – this lexeme is functioning in the majority of languages in the form which is very similar to the original. And its meaning is always quite standard, too. The semantics of information has undergone many modifications in the process of speech practice, but at the same time one can hardly assume with certainty that the semantisation of this notion has been completed, and the final significance has already formed.

Until the middle of the XXth century the meaning of information was very common and vague: 'knowledge', 'facts'. With the appearance of computers its content plane acquired many shades: it became comparable with such fundamental scientific concepts as "matter" and "energy". Information has become an extraordinary multi-pronged concept – leaving behind, for example, banal "data" – and continues to uncover itself wider and deeper. Generally, ***information*** is the meaningful specifics of communication, conceptual-speech quintessence of language units in context.

N. Wiener, father of Cybernetics, explains the essence of information through its *functionality*:

"Information is a name for the content of what is exchanged with the outer world as we adjust to it, and make our adjustment felt upon it" [20].

Without exhausting all the semantics of modern understanding of information, this definition illustrates the *technological* approach to the interpretation of communicational content. The main drawback of this approach is the transfer of the significance from the human relations sphere to supposedly independent "external" world of human communication. It presumes that, according to the proposed logic, the laws that act in technical and human communications are different. The next logical step in this direction could be done towards admitting the need to adapt the human communication to artificial standards. This step is often taken because it seems quite simple. There are many evidences of the attempts to simplify human interaction down to computer-acceptable level in practice. The contrary achievements are not known. They are impossible: *heuristics* is too deep for a too shallow *algorithm*. Still nobody can fully align human and computer-mediated communication. Thus, there are indicative limitations for the significance of the lexeme "information" in the world of algorithm:

"The word information, in this theory [statistical], is used in a special sense that must not be confused with its ordinary usage. In particular, information must not be confused with meaning… To be sure, this word information in communication theory relates not so much to what you do say, as to what you could say" [21].

Meanwhile, such explanations ("facts" or "data") are clearly helpless for realistic representation of information. Moreover, they do not work. Here, alongside with the well-known obvious informational characteristics such as "obtained from investigation", "representing data", "which justifies change in a construct" [22], information additionally gains attributive computer-mediated characteristics such as "circulating in a network environment", "contained in electronic format", "available via the Internet", etc. The modern *conceptualization* of information is inextricably linked with the dominating type of interaction, actualized through it: today information "lives" in the processes of *creating*, *storing* and *transferring* significance.

The communication sphere is the area of a huge amount of information that functions and is being permanently updated. Related mechanisms are still unknown; however it does not prevent the human mind from its successful *processing*. Of course, human mechanisms of information processing have a little to do with duplication; they work correctly and are constantly improved. Today we continue the ancient process of

multi-channel complicated communication, started some time by the first human generations. And the quality of "natural information", that is inherent to human intellect, and "artificial information", that is inherent to computer program, looks very different. So, if the *Natural Information* reflects the contextual potential of the intuitive kind for the heuristic interpretation in natural-language environment, the *Artificial Information* is the algorithmically correct textual add-on for processing of speech in the discrete environment of formal languages.

The complex of essential development in the functional aspect of communication includes the issues of structuring, preserving and processing of information. And computer-mediated communication not only objectively contains information, but makes it more *visible*. Contemporary communication is characterized by the accessibility of empirical material for any large-scale research, providing it with effective technological support. With that, enormous quantity of aggregated speech provides new quality of research, allowing considering the samples of speech under the "hyperscope" of its informational functionality. Such informational "hyperscope" will be as useful as the statistical "telescope", mentioned by T. McEnery and A. Hardie [23]. Of course, future discoveries will depend on it.

No doubt that the effectiveness of computer facilities – including *transfer* of information – is already incomparably superior to the statistical and algorithmic abilities of humans. Step by step the possibilities for describing the content aspects of modern communication are irreversibly mediated by the computer-dependent technologies. But until today superficial potential of such activity is exhausted. Then we want to *create* significance with the assistance of computers. And right here the effectiveness of computers will be conditioned by the depth of meta-language representation of generalized semantics – information. Today no researcher changes the computer keyboard for a typewriter or fountain pen. Tomorrow, with the assistance of Artificial Intellect, the main instrument could be information.

V. Conclusion

So, no one has ever seen either information or language. But as the ephemeral nature of language does not interfere with the existence of Linguistics – in the same way the ephemeral nature of the information is not an excuse to ignore Informational Linguistics. Information is an abstraction, like everything in the mentality world. With that, it has already gained "critical mass" of conceptualization and promotes *de facto*, in turn, the distinct identity acquisition for ***Informational Linguistics*** – a full-fledged discipline, dedicated to the interdisciplinary investigation of the communication content specifics.

Practice shows that the roads to syncretic knowledge are thorny and related specializations are quite new for universities. But at the same time, other "universities" – such as life – are attracting dozens of newcomers into related fields of work. These specialists are different in qualifications, but equal in their thirst and readiness to change the surrounding world and future. So, probably the way, in which to announce the *visualization* of Informational Linguistics, is not essential. Informational Linguistics could be considered as a new generation of "Informatics". It could be considered as a new branch of Linguistics. With that the Informatics – like *Information Technologies* themselves – is always accepted as some abstract *metaphor* for communicational knowledge. On the contrary, Informational Linguistics is more real and actually a whole unit. Obviously, the informational dimension of communication is relevant, first of all, from the linguistic point of view. With that, the point is that such actualization fits the common human field, artificial and somewhen hybrid intellectual interaction.

In the context of these circumstances, Informational Linguistics occupies a *special place*: it is on the cutting edge of scientific and technological progress and completely dissolved in daily life: switching on a computer and connecting to the Internet, we are studying it anyway. The only problem is to study it efficiently with eyes open.

References

[1] M. Foucault, "The archeology of knowledge," London : Tavistock Publications, 1972.

[2] A. Barkovich, "Internet Discourse: Computer-Mediated Communication" [Интернет-дискурс: компьютерно-опосредованная коммуникация], Moscow : Flinta; Science. 2015.

[3] A. Barkovich, "Informational Linguistics: Meta-Descriptions of Modern Communication" [Информационная лингвистика: метаописания современной коммуникации], Moscow: Flinta; Science, 2017.

[4] E. Benveniste, "Problems in General Linguistics," University of Miami Press, 1971.

[5] A. Barkovich, "Corpus Linguistics: the Specifics of Modern Meta-Descriptions of Language" [Корпусная лингвистика: специфика современных метаописаний языка], in Tomsk State University Journal, Tomsk, No. 406, 2016, pp. 5–13.

[6] A. Kornai, "Mathematical Linguistics," London : Springer, 2008.

[7] W3Techs – extensive and reliable web technology surveys, available at: http://w3techs.com.

[8] D. Crystal, "Language and the Internet," Cambridge : Cambridge University Press, 2001.

[9] T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, "A Framework for Web Science," in Foundations and Trends® in Web Science, Vol. 1, No. 1, 2006, pp 1–130.

[10] N. Shadbolt, W. Hall, T. Berners-Lee, "The Semantic Web Revisited," in IEEE Intelligent Systems, Vol. 21, Iss. 3, 2006, pp. 96–101.

[11] A. Barkovich, "Computer-Mediated Communication: Potential of Metalexical Significance" [Компьютерно-опосредованная коммуникация: потенциал металексической значимости], in Proceedings of Petrozavodsk State University, Petrozavodsk, No. 7 (152), 2015, pp. 38–43.

[12] A. Barkovich, "Functionality of "Communicational–Communicative" Dyad: a Discursive Aspect" [Функциональность диады «коммуникационный–коммуникативный»: дискурсивный аспект], in Tomsk State University Journal of Philology, Tomsk, No. 5 (37), pp. 37–52.

[13] M. Minsky, "A framework for representing knowledge," in Haugeland, editor, Mind Design: Philosophy, Psychology, Artificial Intelligence, 1981, P. 95–128.

[14] P.N. Johnson-Laird, "Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness," Harvard University Press, 1983.

[15] R.C. Schank and R.P. Abelson, "Scripts, plans, goals, and understanding : an inquiry into human knowledge structures," Hillsdale, N.J. : Lawrence Erlbaum Associates, 1977.

[16] T.A. van Dijk, "Episodic Models in Discourse Processing," in Comprehending Oral and Written Language, New York : Academic Press, 1987, pp. 161–196.

[17] S.C. Herring, "Computer-mediated discourse," in Handbook of Discourse Analysis, editors D. Tannen, D. Schiffrin and H. Hamilton, Oxford : Blackwell, 2001, pp. 612–634.

[18] C.D. Manning, P. Raghavan and H. Schütze, "Introduction to Information Retrieval," Cambridge : Cambridge University Press, 2008.

[19] C.E. Shannon, “The mathematical theory of communication,” in The Mathematical Theory of Communication, Urbana : The University of Illinois Press, 1964, pp. 29–125.

[20] N. Wiener, “The Human Use of Human Beings : Cybernetics and Society,” Boston : Da Capo Press. 1954.

[21] W. Weaver, “Recent Contributions to the Mathematical Theory of Communication,” in The Mathematical Theory of Communication, Urbana : The University of Illinois Press, 1964, pp. 1–28.

[22] Dictionary by Merriam-Webster : America’s Most Trusted Online Dictionary, available at: https://www.merriam-webster.com.

[23] T. McEnery, A. Hardie, “Corpus Linguistics: Method, Theory and Practice,” Cambridge : Cambridge University Press, 2012.