

Estimate of the distance between areas of an organization using the concept of interlinguistic distance

Miguel Filho Ferreira de Oliveira¹, Marcelo Schiessl²

¹Caixa Econômica Federal, Brasília, Brazil. migueloliveira@terra.com.br.

² Universidade de Brasília, Brasília, Brazil. marcelo.schiessl@gmail.com.

Abstract

The concept of “interlinguistic distance” is used to estimate the distance (linguistic) between two or more areas of an organization. It considers the diversity between lexicons of specialized languages within the organization and the employees familiarity related to the terms of these languages. The organizations are perceived as multilingual communities, endowed with a great variety of languages; “linguistics Babel”, where some specialized languages cohabit to reach common organizational objectives. The larger the level of diversification, the larger the diversity is at the language level, the quantity of specialized language and the difficulty of communication. It analyses the language produced by a group to apprehend different levels of abstraction of the same object or to visualize the outline of relations between generalizations, specialization of organization objects. It aims to present a metric to estimate the linguistic diversity among several groups and areas of an organization. The result is measurements of similarities between the specialized language and the experts which show how each one is far from the specialized languages.

Keywords: Interlinguistic distance, Organizational language, Similarity, Specialized language.

Resumen

Se utiliza el concepto de “distancia interlingüística” para estimar la distancia (lingüística) entre dos o más áreas de una organización. Se considera la diversidad entre lexicones de lenguajes especializados utilizados dentro de una organización y la familiaridad de los empleados con los términos de cada uno de estos lenguajes. Las organizaciones son percibidas como comunidades multilingües, dotadas de una gran variedad de lenguajes, una “babel lingüística” donde cohabitan algunos lenguajes especializados para alcanzar objetivos organizacionales comunes. Cuanto mayor es el nivel de diversificación, mayor es la diversidad en el nivel del lenguaje, la cantidad de lenguaje especializado y la dificultad de

comunicación. Se analiza el lenguaje producido por un grupo con el fin de aprehender diferentes niveles de abstracción del mismo objeto o de visualizar el esquema de las relaciones de generalización y especialización entre los objetos de la organización. Se pretende presentar un sistema de medidas para estimar la diversidad lingüística entre varios grupos y áreas de una organización. El resultado es la medición de la similitud entre el lenguaje especializado y los expertos, mostrando como dichos expertos se encuentran alejados de los lenguajes especializados.

Palabras clave: Distancia interlingüística, Lenguaje especializado, Lenguaje organizacional, Similitud.

1 Introduction

Language in the organizations is a “space” where one can see reflected the inner difference of the organization in a form of specialized languages. The larger the level of diversification, the larger the diversity is at the language level, the quantity of specialized language and the difficulty of communication as well.

The dynamic and specialization of several environments of the organization bring new expressions, new terms, new definitions to old terms, new forms of concordance between verbs and nouns, etc., compelling technicians to acquire better specialization at the language level, which moulds them and is moulded by them.

This phenomenon gives opportunity to the rise of jargon, which differentiates classes of workers by language. Diversification at the language level is a sign of inner differentiation; it is a positive and expected fact in a contingency point of view.

This article aims to present a metric to estimate the linguistic diversity among several groups and sectors of an organization.

2 Division of labour and the languages of the work

The division of labour is a collective strategy used to accomplish a task, to yield goods or to offer services, in which every worker accomplishes part of the global task. It has been done this way since the very beginning of mankind until now, and in practically all aspects of modern life.

In organizations, it is used - almost obligatory – and is justified when one wants to accomplish a task requiring physical efforts or intellectual capacity of more than one person; or when one intends to minimize the cost of the final product or to yield more standardized goods. In spite of the ubiquitous character, its use in organizations does not constitute a panacea: its effective use, in a social-technical perspective of the organization, requires an analysis of several prerequisites and the evaluation of the consequences because of its implementation.

The division of labour can be done based on some criteria or on a composition of them. It can be guided by variables external to the organization, as in the case of a division of labour based on the product, customer or on the geographic localization; or it can follow an internal orientation based on the way the work is executed. The work can also be subdivided into

several levels. In some cases, it can be subdivided into macro tasks, which require the action of generalists in order to be executed, or into specific actions requiring the intervention of specialists.

Independently of the adopted criterion or the level of the subdivision, every division of labour results from the grouping of people of an organization in distinct groups of affinity, or operative groups, being responsible for one or several organizational sub-objectives. This horizontal division is cited by Litterer (1970) as system of operation (SO). The groups present deficiencies, which, if not compensated, can harm the achievement of the global objectives. Its elements can prioritize excessively a local vision in detriment of a global vision, or acquire excessive autonomy leading to the isolation of the remaining portion of the company and causing damage to the final results of the organization. As a direct consequence of the division of labour, the necessity to integrate and to control the operative groups through an administration system (AS) arises (Litterer, 1970). Thus, every division of labour establishes a relationship between two opposing forces: one which divides, and results in the system of operation (SO), and the other (AS) which integrates and controls the SO (Lawrence, 1973).

Many factors concerning the company and the environment – including technology and the availability or lack of specialists in the market – influence the way the labour is divided. However, once it is defined, it will start to exert a great influence on several important aspects of the organization. The division of the adopted labour defines, for example, the dimensions of an organization which will be more or less flexible, the characteristics of the administration system, and even how and with whom a person will have chance to interact within the organizational environment.

An important consequence of the division of labour occurs in the language level, but it is paradoxically still under developed by the administration theoreticians (Oliveira, 1996), (Girin, 1989). The same force derived from the division of labour that separates people and tasks in groups is reflected in an isomorphic way at language level: each group adopts and develops in a distinctive way its own language (language + speech, in the Saussurean sense), (Saussure, 1995), as a result of a complex process of influences including the education level and profession of each element, the tasks in charge of the group, the objectives and the activity branch of the company, the information flow that links the group to the other sectors of the company, the autonomy degree of the group, and the development stage of the technological sector of the company (Oliveira, 1996). This linguistic specificity inherent to each group facilitates many aspects of the group; on the other hand, it makes the integration and the communication among the groups difficult. Thus, the more specific and isolated they are, the more serious are the problems to the objectives of the organization. Lesca (1981) considers this linguistic diversity in organizations an “undesirable situation because of its harmful effect to the communication” and Strassman (1981) proposes a “standardization of the language” or “to invent a more precise language”.

An alternative to be considered is to preserve – or to stimulate – the linguistic plurality of each group and, an example would be the administrative system which integrates the operative groups considering their objectives, to provide the organization with a *management system of the organizational language* in order to manage the language of the organization as an important input and essential product to the organizational dynamics derived from the integration of several specialized languages (Oliveira, 1996). This point of view is based on the ethnolinguistic premise that “the world tends to be perceived by the members of the

organization in terms of the concepts reflected in the vocabulary of the organization” (Simon, 1972). Thus, the bigger the linguistic pluralism, controlled, the better the internal communication of the organization becomes and its perception of the external environment.

This article intends to contribute to this point of view proposing a metric with the aim of estimating the linguistic diversity between several operative groups based on their vocabularies.

3 Language as a mirror of the group

To say that each group adopts a language is to indirectly affirm evidence. However, for the purpose of this paper, which consists in “estimating the distance between organizational groups based on their languages”, what is of interest is: given there are no two equal groups (even considering the same group in two different moments), by comparing their characteristics, are their specificities reflected in the language of the group? That is, in which measure, analyzing two “parts” of the language (called here as “corpus”) C1 and C2, “conveniently chosen” from groups G1 and G2 respectively, can we conclude that they had come from different groups?

Studying this question, Oliveira (1996) concluded that many characteristics of the organization and the groups are reflected in the language, and this relation is mediated by the language, since the organizational dynamics gains dimension at the linguistic level, but this imposes “use of rules” of morphologic, syntactic, semantic, and pragmatic nature in order to preserve the language as a communication instrument. The first result obtained by Oliveira (1996) is related to the aspects of technology that are reflected in the language.

Certain kinds of companies are inserted into markets, strongly based in technology; it is the case of sectors such as chemicals, electronics, automobiles, etc. According to Lawrence (1973), these companies are facing increasing development in terms of scientific knowledge, and the competitiveness of each one of them is placed in their innovation capacity either in terms of methods or products. The process of naming a new product is guided by rules of the mother language (Portuguese, Spanish, French, English, etc.) to which the group language belong.

Every mother language makes use of mechanisms in order to form new words. In the case of the Portuguese language they are: prefixal, suffixal, parasynthetical, regressive and improper derivation, composition by juxtaposition and agglutination, vocabulary abbreviation, acronyms and onomatopoeias. The elaborated term may have the form of a simple or composed noun, a noun phrase, an acronym, etc., with the function to nominate and to describe things and objects. Through the same process, new verbs or noun phrases may appear which are appropriate to describe methods.

The innovation corresponds – to the language level – to the appearance or disappearance of terms in the organization language or to the change of meaning associated to the terms. The double verb-name gives an idea of the processes that are applied to objects, or what objects receive the action represented by a verb.

Some aspects of the coordination are also reflected in the language. According to Mintzberg (1982), coordination “is the glue that keeps joined parts of the organization” and classifies it

in five different types: a) the mutual adjustment; b) direct supervision; c) standardization of procedures; d) standardization of products; and e) standardization of qualifications. The mutual adjustment accomplishes the work coordination through the “informal communication” (Mintzberg, 1982). The control of the work is in charge of the interlocutors. According to this author, due to its simplicity, this modality at the same time is used in the simplest situations and paradoxically in the most complex ones. The forms of coordination, mutual adjustment and direct supervision are, in general, expressed in the form of language not differed. The coordination mechanisms by standardization (procedures, products and qualifications) have linguistic equivalents of the differed type.

Various aspects can be found, by analyzing the language produced by a group, one can apprehend different levels of abstraction of the same object or to visualize the outline of relations between generalizations, specialization of organization objects.

In this case, the demonstrative pronouns and adjectives are the elements of the language that establish these relations, for example: “the entrance fees in FINANCIAL SECURITY, for the COLLABORATORS, are 1% instead of 3%. This new contract of life insurance allows guaranteeing the DUPLICATION OF THE CAPITAL invested to the end of 10 years”. We conclude that FINANCIAL SECURITY is a life insurance contract.

The objects and people that are agents of the organizational dynamics and evidences of the relation and cooperation among them can also be observed in the plan of the language through some prepositions, in the same way as the relations between the interlocutors and objects of the speech. In this case, the possessive adjective elaborates an idea of hierarchy between the sender and the receiver; and the degree of sharing of missions, facts, events; indicating “what” is in charge of whom.

4 Metrics of interlinguistic distance

Measuring similarities and differences between two texts is an issue that interests several disciplines (Rajman, 2006). This issue is mentioned in the bibliography concerning the field of lexical statistics for the terms “lexical connection” or “lexical distance” and it consists of measuring the distance existing between (among*) vocabularies of the texts to be compared. In the documentary research, the evaluation of similarity between documents is used to identify documents that are pertinent to the information needs of users. In the analysis of textual data, the similarities are used to describe and explore data, and to exploit hidden structures as well. In text mining field, similarities are used to produce synthetic representations from vast document collections in the process of extracting information from textual data.

However, the recognized application, and also controversial, consists in measuring the similarity between texts that are supposed to be written by an author and other texts of a well-known authorship. Then, one concludes whether the text belongs to the same author or not.

4.1 Jaccard’s Metric

Some metrics are used to calculate distances, for example, the metric of Jaccard, Minkoswki, Salton, Muller, Mahalanobis, Kolmogorof, Hamming, Labbé and others (Brunet, 2003). The evaluation of their properties compared to the characteristics of the phenomenon in study indicated that the most appropriate metric to represent it would be Jaccard metric; but

this, according to Brunet (2003), could not be applied directly because of the two factors explained below.

Jaccard's metric is an Euclidean's metric based on a very simple principle. It establishes that the distance between two texts is the proportion between the amount of words that belong solely to a text and the amount of words from both texts. In operative terms, the distance of Jaccard between two texts A and B, as cited in Brunet (2003) and denoted here by $J(A,B)$, is defined by the addition of two parcels: first, equal the amount of terms that belong only to A, denoted by $n(A-B)$, divided for the amount of terms of A, denoted by $n(A)$; and the second, equals the amount of terms that belong only to B, denoted by $n(B-A)$, divided for the amount of terms of B, denoted by $n(B)$.

$$J(A,B)=n(A-B)/n(A) + n(B-A)/n(B)$$

The Jaccard distance is a real number $J(A,B) \in [0;2]$, $n(A)$; and satisfies the four axioms of a measure of distance:

- 1) Axiom: $d(a,b) \geq 0$;
- 2) Axiom: $d(a,b) = 0$, if and only if $x = y$;
- 3) Axiom: symmetry: $d(a,b) = d(b,a)$;
- 4) Axiom: triangular inequality: $d(a,c) \leq d(a,b) + d(b,c)$.

Two factors have denied its direct application:

The first one is that Jaccard's metric is symmetrical, which is equivalent to say that given two texts A and B, the distance of A "to" B is the same of B "to" A, which is, the direction not considered. It seems to be sufficiently obvious when one deals with physical measures in the Euclidean plan, but not dealing with signs composed by signified and signifier. An e.g. would be two sectors of a hospital, the medical and nursing sectors. If we could compile the whole lexicon of both of them, we would conclude that the distance from the lexicon of medical sector in relation to the nursing sector would be shorter (in the semantic direction and perhaps in the quantitative direction) than the distance from the lexicon of nursing sector in relation to the medical sector. What should be said about the distance between the lexicon of a financial auditing and an accounting sector? A metric which results in the same value in both directions certainly would not be representing the phenomenon described here.

Another factor within Jaccard's metric is the amount of terms common to both sets of terms are not considered in any particular way. The value is inserted in to the calculation in an indirect way associated to the quotient' denominator. A possible interpretation of vocabulary common to both sectors of the organization is that it represents a convergence region between the two languages of the sectors in focus.

4.2 The proposed metric

Considering the two sectors S1 and S2 of an organization with their languages L1 and L2. From each one of them, one takes out one (or more) corpora (portion of the representative language of the language of the sector), C1 and C2, and from these corpora if one extracts the vocabularies V1 and V2 of the sectors S1 and S2. Let us admit that every term of V1 and V2 has the same semantic weight and that the distance between the vocabularies V1 and V2

represents the distance between the languages L1 and L2, or the (linguistic) distance between sectors S1 and S2.

The metric M proposal is calculated on the basis of the following concepts:

- a) DISTANCE FROM V1 TO V2 – The distance from V1 to V2, denoted for $M(V1, V2)$, is what lacks in V1 so that it equals to V2. In operational terms it is the amount of terms that belong exclusively to V2, denoted by $n(V2)$, divided by the amount of elements of V1 plus the amount of elements of V2, that is $n(V1) + n(V2)$. In operative terms, we have:

$$M(V1, V2) = n(V2 - V1) / [n(V1) + n(V2)]$$

- b) DISTANCE FROM V2 TO V1 – In analogous way, the distance from V2 to V1 is defined by:

$$M(V2, V1) = n(V1 - V2) / [n(V1) + n(V2)]$$

- c) BOUNDS BETWEEN V1 and V2 (or V2 and V1) - The shared terms (intersection) V1 and V2, represent a kind of “convergence” between V1 and V2, related here as “bound” between V1 and V2 and denoted by $F(V1, V2)$. In operational terms, the bound between V1 and V2 (or V2 and V1) is:

$$F(V1, V2) = 2 [n(V1 \cap V2) / (n(V1) + n(V2))]$$

Then, the metric M satisfies the following properties:

- 1) $M(V1, V2) + M(V2, V1) + F(V1, V2) = 1$
- 2) $M(V1, V2) \geq 0$, $M(V2, V1) \geq 0$ e $F(V1, V2) \geq 0$;
- 3) $M(V1, V2) = 0$, if and only if $V1 = V2$;

Just to illustrate, it considers V1 to be composed by 440 terms, V2 by 250 terms and 85 terms are common to V1 and V2 and every term has the same semantic value. In this case, the distance from V1 to V2 is $M(V1, V2)=0,2391$, the distance from V2 to V1 is $M(V2, V1)=0,5145$ and the bound between V1 and V2 is $F(V1, V2)=0,2463$.

5 Conclusion

The metric M is being applied to a large banking institution of Brazil, with significant results in diverse sectors. The administrators have an available indicator that allows them to find bottlenecks in the communication of the company, to assemble more homogeneous teams, to gather people who speak similar languages, to plan people training, to elaborate strategies to approximate sectors and people, and to identify similar sectors based on its language. The employees have available and useful information to plan their careers and a better understanding of the languages used in the organizations.

References

BACHELET, C. ; MOSCAROLA, J. *La communication interne: diagnostic par l'analyse du discours des acteurs*. Savoie: Université de Savoie (GEREG), 1995.

- BORBA, F. S. *Introdução aos estudos lingüísticos*. São Paulo: Companhia Editora Nacional, 1972.
- BRAM, J. *Linguagem e sociedade*. Rio de Janeiro: Bloch, 1968.
- BRUNET, E. Peut-on mesurer la distance entre deux textes? *Revue CORPUS*, 2003, n. 2.
- CROZIER, M. Ce que nous a apporté Herbert Simon. *Revue Française de Gestion*, 1993, n. 94, p. 85-86.
- DRUCKER, P. L'information dont un dirigeant a vraiment besoin. *L'Expansion Management Review*, 1995, p. 6-14.
- FOUCAULT, M. *Les mots et les choses*. Paris: Gallimard, 1966.
- GIRIN, J. (ed.). Séminaire CONDOR (Paris: 1989). *Contradictions et dynamique des organizations: actes*.
- KENNEDY, C. *Toutes les théories du Management*. Paris: Maxima, 1993.
- KLUCKHOHN, C. *Antropologia, um espelho para o homem*. Belo Horizonte: Itatiaia, 1963.
- LABBÉ, C.; LABBÉ, E. La distance intertextuelle. *Revue CORPUS*, 2003, n. 2.
- LABBÉ, D. *Corneille et Molière*. Paris: Université Paris XI-Orsay, 2004
- LABBÉ, D.; HUBERT, P. La connexion dès vocabulaires. In: *Sommaire dès JADT*, 1998
- LÉVI-STRAUSS, C. *Anthropologie structurale*. Paris: Librairie Plon, 1974.
- LITTERER, J. A. *Análise das organizações*. São Paulo: Editora Atlas, 1970.
- LORINO, P. *Le contrôle de gestion stratégique: la gestion par les activités*. Paris: Dunod, 1991.
- LUSSATO, B. *Introduction aux théories d'organisation*. Paris: Dunod, 1992.
- MINTZBERG, H. *Structure & dynamique des organisations*. Paris: Editions d'Organisations, 1982.
- MOSCAROLA, J. *Manuais do sistema "Le Sphinx Lexica"*, 1996.
- MOSCAROLA, J. *L'analyse de contenu et analyse de données assistées par ordinateur, nouveaux outils et nouvelles pistes: le projet ISIS*. Savoie: Université de Savoie (GEREG), 1994.
- MOSCAROLA, J. From the utterance to the enunciation for a re-reading of lexical analysis in marketing. *International Research Seminar on Marketing Communication and Consumer Behavior (1st: Lalond les Maures: 1995)*.
- OLIVEIRA, M. F. F. de. *Le diagnostic et l'action sur les organisations à travers la représentation du langage : thèse de Doctorat ès Sciences de Gestion*. Savoie: Université de Savoie, 1996.
- PETLAND, B. Grammatical models of organizational processes. *Organization Science*, 2000.
- PETLAND, B [et al.]. Lexical and sequential variety in organizational processes: some preliminary findings and propositions. *Organization Science*, 2000.
- PETLAND, B. A grammatical models of organizational routines. *Administrative Science Quartely*, 2000.
- RAJMAN, M.; LEBART, L. *Similarités pour données textuelles*.

SAPIR, E. *Linguistique*. Paris: Gallimard, 1968.

SAUSSURE, F. de. *Cours de Linguistique Générale*. Paris: Payot, 1995.

SIMON, H. A. *Les Organisations*. Paris: Dunod, 1969.

SIMON, H. A.; MARCH, J. G. *Teoria das Organizações*. Fundação Getúlio Vargas, 1972.

WARLICH, B. M. des. *Uma análise das teorias de Organização*. Rio de Janeiro: FGV.