Texto Livre: Linguagem e Tecnologia
E-ISSN: 1983-3652
revista@textolivre.org
Universidade Federal de Minas Gerais
Brasil

Blanch Pires, Thiago
Multimodality and evaluation of machine translation: a proposal for investigating intersemiotic mismatches generated by the use of machine translation in multimodal documents
Texto Livre: Linguagem e Tecnologia, vol. 11, núm. 1, january-april, 2018, pp. 82-102
Universidade Federal de Minas Gerais

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

MULTIMODALITY AND EVALUATION OF MACHINE TRANSLATION: A PROPOSAL FOR INVESTIGATING INTERSEMIOTIC MISMATCHES GENERATED BY THE USE OF MACHINE TRANSLATION IN MULTIMODAL DOCUMENTS

MULTIMODALIDADE E AVALIAÇÃO DE TRADUÇÃO AUTOMÁTICA: UMA PROPOSTA PARA A INVESTIGAÇÃO DE INCOMPATIBILIDADES INTERSEMIÓTICAS GERADAS PELO USO DO TRADUTOR AUTOMÁTICO EM DOCUMENTOS MULTIMODAIS

Thiago Blanch Pires
Universidade de Brasília
thiagocomaga@gmail.com

ABSTRACT: This article aims at proposing an interdisciplinary approach involving the areas of Multimodality and Evaluation of Machine Translation to explore new configurations of text-image semantic relations generated by machine translation results. The methodology consists of a brief contextualization of the research problem, followed by the presentation and study of concepts and possibilities of Multimodality and Evaluation of Machine Translation, with an emphasis on the notion of intersemiotic texture, proposed by Liu and O'Halloran (2009), and a study of machine translation error classification, proposed by Vilar et. al. (2006). Finally, the article suggests some potentialities and limitations when combining the application of both areas of investigation.
KEYWORDS: multimodality; machine translation; evaluation of machine translation; intersemiotic mismatches.

RESUMO: Este artigo tem como objetivo propor uma abordagem interdisciplinar envolvendo as áreas da multimodalidade e da avaliação de tradução automática para explorar novas configurações de relações semânticas entre texto e imagem geradas por resultados de traduções automáticas. A metodologia é composta de uma breve contextualização sobre o problema de investigação, seguida da apresentação e do estudo de conceitos e possibilidades da multimodalidade e da avaliação de tradução automática, com destaque para os trabalhos respectivamente sobre textura intersemiótica proposta por Liu e O'Halloran (2009) e classificação de erros de máquinas de tradução proposta por Vilar et. al. (2006). Ao final, o estudo sugere algumas potencialidades e limitações no uso conjugado de ambas as áreas.
PALAVRAS-CHAVE: multimodalidade; tradução automática; avaliação de tradução automática; incompatibilidades intersemióticas.

1 Introduction

Websites of various contents are multimodal documents often required for online automatic translation systems. But what part of this automated translation is exactly considered information may be at least part of a "combination of different modes of information" (BATEMAN, 2008) generated in a form displayed for the user. In other words,

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

the way visual and verbal components of a text are combined may reveal meaning potential to be automatically translated.

The communication and socio-semiotics interdisciplinary approach informing multimodality (KRESS; VAN LEEUWEN, 1996, 2001) has in the last decades developed a growing number of works about communicative practices that use visual, verbal, auditory, and spatial resources (called “modes”) to compose messages.

Although such descriptions and analytical categories in text-image relation have expanded, there is still little investigation on such relationships within the context of machine translation output. A lack of investigation is also observed in the area of Computational Linguistics, specifically manual evaluation of machine translation which studies this relationship from machine translation error typologies.

The use of the term “errors” to refer to translation within the context of computer science is generally used in the mathematical sense (i.e. in the sense of calculation). For that reason, “errors” are rarely conceptualized or discussed when they are applied to the use of machine translation.

Thus, the present paper draws on the social semiotics perspective (multimodality) and manual evaluation of machine translation about such “errors” within a given production context (such as webpages, illustrated manuals and infographics) to propose a context of investigation to analyze linguistic errors (lexical, semantic and syntactic) between the input text and the output text generated by machine translation (MT).

First, it discusses the research problem by describing an example of part of a BBC article automatically translated by means of Google Translate from English into Portuguese. Then, it explains and explores some key concepts of Multimodality and Evaluation of Machine Translation, highlighting Liu and O’Halloran’s (2009) “intersemiotic texture” and Vilar et. al.’s (2006) machine translation error types. Finally, it suggests potential pathways for joining both areas to explore the problem of text-image relationships in automatically translated multimodal documents.

2 What’s the real problem?

Introducing the research problem and contextualizing it may not be entirely sufficient. Perhaps showing some images of the phenomenon under investigation could help to visualize the problem.

The following images were captured from a BBC news article, which revolves around Islamic state militants destroying ancient history in Syria[1] in which there is a video of a man pushing a statue. The first image below shows a screenshot of the video and its caption originally in English; the subsequent image shows the same part of the article translated with the Google Translate add-on into Portuguese.

1 To access the article, refer to <http://www.bbc.com/news/world-middle-east-32820857>.

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

Islamic State (IS) militants in Syria have entered

Estado Islâmico (IS) militantes na Síria entraram

Figures 1 and 2: Screenshots of video and caption from BBC online news article in English and its automatic translation into Portuguese.
Source: BBC news

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

The problem here starts with the reading of IS (Islamic State) as the verb “is”. This confuses the meaning of the caption in Portuguese, making it partially nonsensical. However, when this traditional machine translation issue is viewed via a text-image relationship, we can also detect that the video loses the verbal attribution of IS (Islamic State) as it is in the original caption.

Such new configurations of text-image semantic relationships, triggered by an automated translation, may generate new meanings or even compromise the reader’s comprehension (depending on languages, text genre, level of reading skills).

Machine translation systems are not designed to detect intersemiotic relationships, but if there is a pattern of these connections, then they must be formalized to improve these systems’ precision.

Therefore, there are two approaches that could be used to support detecting text-image relationships. One that serves as a resource for detecting semantic relations between image and text, and another that serves as a basis for manually classifying certain linguistic errors generated by machine translation. These two types of resources are typically used by Multimodality and Evaluation of Machine Translation.

3 What can multimodality do?

3.1 General concepts

Multimodality is an interdisciplinary approach based on communication theory and social semiotics that moves towards a theory (KRESS; VAN LEEUWEN, 2006). Language beyond language, according to some scholars, multimodality is concerned with comprehension and representation, isolated and interconnected, of nonverbal forms of communication that people use, such as gesture, posture, image, and other forms (JEWITT, 2009).

Jewitt (2009, p. 14-16) defines four assumptions that support a concept generally shared among multimodality scholars:

> • Language is part of a multimodal set. This means that multimodality understands that representation and communication always influence a multiplicity of modes (such as gestures, postures, images and sounds), all with potential to equally contribute with a meaning;
> • Each mode of a multimodal set construes different communicative works. Multimodality assumes that, as with language, all modes are shaped by means of their social, cultural, historical use to construe social functions. Therefore, images and other non-linguistic modes have a given role, within a given context and in a given moment;
> • Individuals orchestrate meaning by means of the selection and configuration of modes. Thus, the interaction among modes is important to make meaning. Meanings in any mode are always interconnected with meanings made with those from other modes, which are co-present and co-operative in the communicative event;
> • Meanings of signs shaped from multimodal semiotic resources are, such as speech, social. That is, they are shaped by norms and rules that operate at the moment a sign is construed, influenced by motivation and interests of the signaler

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

> in a given social context. In other words, sign providers select, adapt and re-shape meanings by means of the reading/interpreting process of the sign.

The assumptions described by Jewitt (2009) represent a convergence of some authors, though there are other relationships concerning the use of mode, expanding or modifying its meaning depending on its context of use. This is highlighted by Pires and Duque (PIRES; DUQUE, 2015, author's translation) as follows:

> [...] the understanding of mode, semiotic resources, modality are articulated in a given context of occurrence so to observe its articulation and manifestation in a context socially and culturally construed. For instance, Kress (KRESS, 2009), p. 54) refers to multimodality as mapping a niche of investigation, given the singularities when applied in different areas, with different problems such as medicine, anthropology, and education, for example. In this sense, one may notice that multimodality follows the scientific development with views of a complex reality in a world increasingly globalized, where different semiotic modes are used to disseminate messages and articulate different media in making meaning with a potential of understanding that extrapolates geographical barriers. Thus, understanding how language technologies are used to achieve such objectives is relevant to the study of information in different languages and cultures.

Kaltenbacher (2004) reviews works that contributed to the elaboration of multimodality as a research area. He investigates the first attempts to analyze different connections between semiotic modes by German classicists, and then studied works from systemic functional linguistics and discourse analysis that contributed to establishing multimodality as a new area of study.

In his work (KALTENBACHER, 2004), he also describes the main areas that supported the multimodal approach as it has been known for the last two decades. The linguistic theory comes partly from the concepts of Halliday's social semiotics (HALLIDAY, 1975, 1978), who is notably the Systemic Functional Linguistics pioneer scholar[2], and who substantially influenced the establishment of multimodality as recently known and recognized as such by the pioneering work of Kress and van Leewen (1996) and O'Toole (1994).

Kress and van Leeuwen (2001), attempt to provide a "common terminology for all semiotic modes" (p. 1), but differently from their previous work in 1996 when they had tried to put together a grammar for the visual, and thus focusing only on one type of mode (visual modes such as images), in this book (KRESS; VAN LEEUWEN, 2001, p. 1) they applied the "common terminology for all semiotic modes [to] a given social-cultural domain, [thus] the 'same' meanings can often be expressed in different semiotic modes".

This new concept sets aside the idea of a fixed specialist role for each mode, such as music is only to be interpreted in terms of sounds, emotion, and so on, defined as "common semiotic principles operating in and across different modes" (KRESS; VAN LEEUWEN, 2001, p. 2) so as to "be possible for music to encode action, or images to encode emotions" (ibid., p. 2).

For Kress and Van Leeuwen (2001, p. 2), "In the age of digitization, the different

2 For more details, see Halliday, (1975, 1978), Iedema (IEDEMA, 2003), Halliday and Hasan (1976) and De Beaugrande and Dressler (1981) respectively.

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

modes have technically become the same at some level of representation, and they can be operated by one multi-skilled person, using one interface, one mode of physical manipulation, 'shall I express this with the sound of music', shall I say this visually or verbally?".

One can observe that with such a perspective, the "unifying and unified" element of technology and semiotics tends towards a multimodal discourse, because for a discourse to happen it should contain cohesive elements that "hang together" to form a coherent and unified idea via interaction of different semiotic modes in a given social practice.

Traditional Linguistics works with the idea of "double articulation", that is when texts are "articulated" as form and as a meaning. Differently, articulate meanings according to Kress and van Leeuwen's (2001, p. 2) multimodal text multiply. In addition, these meanings are articulated in four domains of practice called *strata* (adapted from Hallidayan functional linguistics). These four strata are discourse, design, production, and distribution (ibid., p. 2).

For Kress and van Leeuwen (2001, p. 3) discourse is "a socially constructed knowledge of (some aspect of) reality". They explain that "socially constructed" occurs in very "specific contexts", and in a fashion that it is "appropriate to the interest of social actors in these contexts, which can be broad such as Western culture, or narrow, such as a conversation between siblings (my examples).

Another perspective of multimodality that develops multimodal discourse, though within a more empirical perspective, is John Bateman's work. In "Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents" (BATEMAN, 2008), Bateman elaborates a consistent methodology to empirically analyze multimodal documents, where there is "interaction" and "combination" of multiple modes within a single artifact.

Bateman (2008, p. 1) describes "mode" as all diverse visual aspects in which information is presented. Thus, he explains that:

> Combining these modes within a single artefact—in the case of print, by binding, stapling, or folding or, for online media, by 'linking' with varieties of hyperlinks—brings our main object of study to life: the multimodal document.

Besides this definition of mode, another element that is essential to Bateman's methodological approach is the notion of genre. The scholar (Bateman, 2008, pp.9-11) attributes descriptions that support the notions of genre used in the analysis of multimodal documents, namely: i) the informal notion of genre such as "websites" and "newspapers" and the meaning where these genres are realized; ii) genre allows theorizing about a range of possibilities open to the documents[3]; and iii) to consider the materiality of multimodal artifacts (documents included) as a crucial component in conceiving multimodal genre.

3 Bateman, based on Lemke (1999), Fairclough (1992), Bazerman (1994), characterizes genre "not as a loose collection of separated text types, but as 'points', or better regions, in an entire space of genre possibilities […][since] genres can change, and can hybridise with, and colonise, one another" (2008, p. 10).

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

Bateman (2008) develops a systematic and analytical resource (originally called "GeM Project") of multimodal documents based on "layers" called genre, navigation, layout, and rhetorical structure. Table 1 describes these layers, used to identify the levels of interaction and combination of different aspects for further manipulation (by means of computer programming):

Table 1: Bateman's Genre and Multimodality framework main layers (2008).

Content structure	the content-related structure of the information to be communicated – including propositional content
Genre structure	the individual stages or phrases defined for a given genre: i.e., how the delivery of the content proceeds through particular stages of activity
Rhetorical structure	the rhetorical relationships between content elements: i.e., how the content is 'argued', divided into main material and supporting material, and structured rhetorically
Linguistic structure	the linguistic details of any verbal elements that are used to realize the layout elements of the page/document
Layout structure	the nature, appearance and position of communicative elements on the page, and their hierarchical inter-relationships
Navigation structure	the ways in which the intended mode(s) of consumption of the document is/are supported: this includes all elements on a page that serve to direct or assist the reader's consumption of the document.

Source: Based on Bateman (2008, p. 19).

Table 1 renders the layers and corresponding definitions that inform the Genre and Multimodality Project for analyzing multimodal documents systematically.

3.2 Text-image relationships

The previous subsection briefly explained in a general sense what is multimodality, and showed some foundations and key concepts. This subsection briefly discusses some works within multimodality that develop relationships between text and image that can be used to analyze the research problem explored here.

Bateman (2014) offers some studies within the field of multimodality that explore the aspect of multiplication of senses by means of how visual and textual modes are combined. He (2014, p. 8-10) questions the "natural" view that text and image are two completely distinct components, supported with examples such as the representation of organic compounds and maps.

This is the starting point from which Bateman (2014) demonstrates diverse problems and approaches that deal with such phenomena within multimodality. Among these approaches, the most paramount for the present study's purposes is the "intersemiotic texture" (Ibid., p. 171) within the multimodal cohesion and text-image relationships, which are part of the "linguistic-system based approaches" module.

In said section, the most significant work within the aspects of intersemiotic texture,

Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

according to Bateman (2014), is the study by Liu and O'Halloran (2009). According to Bateman, (ibid., p. 171), the substantiality of this work is given by the expansion of Royce's intersemiotic complementarity concepts (ROYCE, 1998, 2007) and intermodal semiosis found on the mathematical discourse studied by O'Halloran (2005)[4], aiming at offering more of a model to join different modalities, rather than only a documentation of superficial relations.

Liu and O'Halloran's (2009, p. 367) work entitled "Intersemiotic Texture: Analyzing Cohesive Devices between Language" proposes an "intersemiotic texture" as crucial property of coherent multimodal texts, and presents a preliminary model for cohesive mechanisms between language and images (BATEMAN, 2014). Based on Halliday and Hasan's (1976, p. 1-2) idea that "texture" involves the relationships of meanings and constitutes crucial elements of a linguistic text, Liu and O'Halloran (2009, p. 369) add the term "intersemiotic" to treat semantic relations between text and image represented by intersemiotic cohesive elements in a multimodal discourse.

It is worth mentioning that the concept of "multimodality" researched in my study is presented in a general sense, referring to the area of investigation, rather than to its relationship between the different modes of communication; therefore, the present study is similar to the distinction made by Liu and O'Halloran (2009) when using the term "multisemiotic".

According to O'Halloran (2005, p. 20-21) "multisemiotic" refers to texts that use more than one Semiotic resource, i.e. they use more than one medium to make meaning, and "multimodal" is used to denote texts that involve more than one channel of semiosis[5] (for instance, visual, auditory and tactile).

Thus, the authors (LIU; O'HALLORAN, 2009) present a preliminary attempt to categorize intersemiotic texture in a multisemiotic text according to Figure 3:

4 Both concepts of "intersemiotic complementarity" (ROYCE, 1998; 2007) and "intermodal semiosis" (O'HALLORAN, 2005) are not explored here.

5 The term refers to "acts of meaning through choices from language and other sign systems" (O'HALLORAN, 2005, p. 3).

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

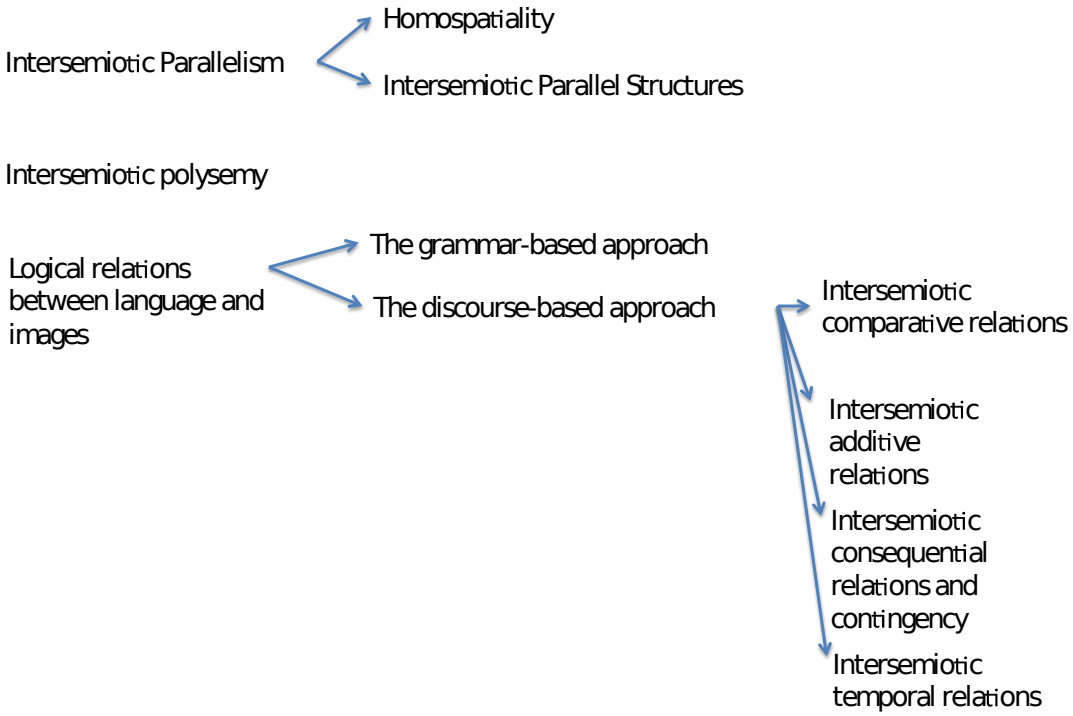

Figure 3: Intersemiotic texture categories proposed in Liu and O'Halloran (2009).
Source: Liu and O'Halloran (2009).

The model proposed by Liu and O'Halloran (2009, p. 372-374) is composed of three main categories called "intersemiotic parallelism", "intersemiotic polysemy", and "logical relations between language and images". The first category, "intersemiotic parallelism", occurs by means of a cohesive relationship that interconnects language and image when both semiotic components share a similar form. This parallelism is formed by "homospatiality" or by "intersemiotic parallel structures". The former is characterized by the parallelism between language and images on the expression plane; and the latter is characterized when language and image share a similar "transitivity"[6] configuration. To illustrate this subcategory, the authors use the following image (Figure 4):

6 Grammatical system composed by types of "processes" (described by verbs), participants of the process, and circumstances related to the processes (cf. HALLIDAY, 1985, 1994; HALLIDAY; MATTHIESSEN, 2004).

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

Figure 4: Example of intersemiotic parallel structure.
Source: Liu and O'Halloran (2009, p. 373).

According to Liu and O'Halloran (2009, p. 373-374), the previous image portrays an action of a dog biting a Muslim woman. This action, represented by a material (physical action according to the transitivity grammar system) is also shared in the description of the caption "Israeli army dog attacks Palestinian woman". Such relation, therefore, can be characterized as an intersemiotic parallel structure.

Another category described by Liu and O'Halloran (2009, p. 375) is "intersemiotic polysemy". In this category, the cohesive relation among the verbal and visual components share multiple meanings in multisemiotic texts. In addition, this type of relation shares similar meanings in opposition to different meanings, generating what some authors call "co-contextualization relations" (ibid. 375). To illustrate this category, Liu and O'Halloran (2009, p. 375) employ the following image (Figure 5):

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

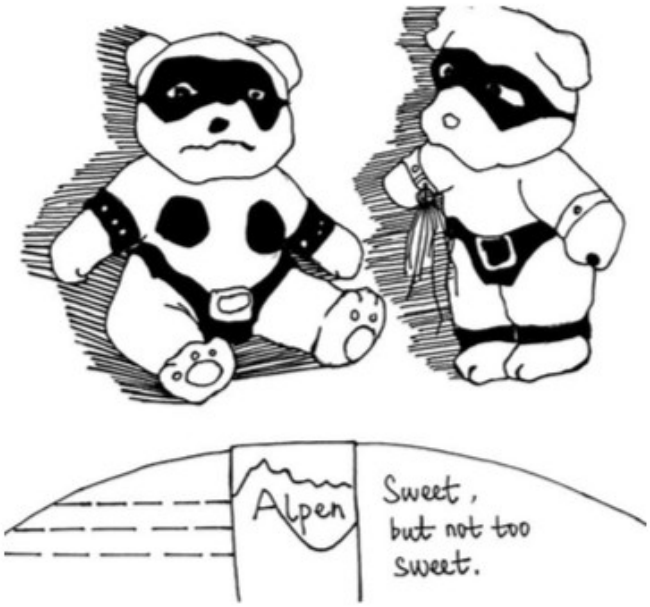

Figure 5: Example of intersemiotic polysemy.
Source: Liu and O'Halloran (2009, p. 375).

The previous image shows an advertisement for Alpen, which is a cereal. The image shows a relationship of meanings between the words "Sweet, but not too sweet", together with the reading of two teddy bears using sado-masochist accessories, generating a polysemous result that is attributed to the cereal brand (LIU e O'HALLORAN, 2009, p. 376-377). Besides that, for the authors (Ibid., p. 375) this text-image relationship uses other intersemiotic relations that cooperate with intersemiotic polysemy, such as "intersemiotic ellipsis" (LEMKE, 1998) and "intersemiotic correspondence" (JONES, 2006).

Intersemiotic correspondence, which differs from a synonym or repetition, characterizes the relationship between a linguistic element and a visual element. In addition, it refers to the joint use among verbal and visual meanings aiming at a resulting meaning correspondence and expansion (JONES, 2006, p. 194).

Differently, intersemiotic ellipsis (O'HALLORAN, 2005 based in LEMKE, 1998) happens when an image or part of it is created to compensate for the lack of grammatical constructions such as the resource "table" in a textualized visual presentation (LEMKE, 1998).

With regards to the third and major category on intersemiotic texture, Liu and O'Halloran (2009) explore logical relationships between language and images, that is, the analysis of logical meanings between verbal and visual components based on grammar and discourse.

According to Liu and O'Halloran (2009, p. 377, based on MARTINEC; SALWAY, 2005) logical relationships between language and images based on grammar[7]:

> [...] provide a preliminary account of the logical meaning across different semiotic resources in old and new media in which language and images are considered either equal or unequal to each other in terms of relative status while the

7 Grammar-based approach draws on Martinec and Salway (2005).

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

intersemiotic logico-semantic relations of projection or expansion apply.

In that passage, the authors (LIU; O'HALLORAN, 2009) explain the analytical limitation of the logical-semantic relationships between language and images, and thus present the need to expand and add such relationships based on grammar and discourse. (based on O'HALLORAN, 2005).

The four subcategories underlying the discourse-based approach are summarized in Table 2:

Table 2: Intersemiotic logical relations and meanings.

Logical Relations	Meaning
Comparative → Generality	Similarity
Comparative → Abstraction	
Additive	Addition
Consequential → Consequence	Cause
Consequential → Contingency	Purpose
Temporal	Successive

Source: Liu and O'Halloran (2009, p. 384).

Table 2 above illustrates four types of discourse-based logical relationships and their respective meanings. According to Liu and O'Halloran (2009, p. 379) the "comparative" intersemiotic relationship is a type of resource used to organized the logical meaning in relation to similarity between the linguistic and visual components in the multimodal discourse, semiotically reformulating one another. Such reformulations are realized on the "generality level" (for example, when the logical-semantic relationship between the linguistic and visual components is realized by means of the general-specific relationship), and in "abstraction" (in cases where the logical-semantic reformulation between part of the visual and linguistic components, the concrete-abstract relationship occurs). The "additive" relationship occurs when a semiotic component adds new information to another semiotic component. The "consequential" relationship can be identified when a semiotic message is perceived as "enabling" or "determining" the other message instead of just preceding it (MARTIN, 1992, p.193, apud. O'HALLORAN, 2009, p. 380). According to Liu and O'Halloran (2009, p. 380), consequential intersemiotic relations can be sub-classified as "consequence" and "contingency", in which the former refers to

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

non-modalized causal relations between verbal and visual messages, where the effect was ensured; while the latter refers to multisemiotic texts where the cause has only the potential to determine the possibility, though without any effect ensured. The fourth category is classified as “temporal” for the steps (sequences) of a procedure represented verbally and visually (generally found in manuals and illustrated guides (IEDEMA, 2003, apud LIU e O'HALLORAN, 2009, p. 383).

As this section has attempted to show, multimodality can offer a substantial number of concepts and approaches for those interested in looking at text-image relationships. More specifically, the second part of this section provided a brief study on some categories that can be employed and expanded in a systematic way.

The next section provides a brief description of the evaluation of machine translation and its possibilities to investigate the phenomenon described in the present study.

4 What can evaluation of machine translation do?

In the context of the studies in the Evaluation of Machine Translation (EMT), there are two main methodologies: one that examines a translation system by looking at its engine, which is called “glass box”; and another which allows the analyst to access only the input and output of a machine translation (MT), called “black box” evaluation.

White (2003) gathers this information to demonstrate forms of evaluating machine translation. As the perspective taken into consideration in this paper looks at the intersemiotic phenomenon emerging from the output of machine translation, the evaluation methodology explored here is the so-called “black box”.

In his work, White (2003, p. 225) uses the following examples to illustrate the benefits of “measure[ing] the coverage of this system, and [to] even have a hypothesis about how the system tries to handle these phenomena”:

Table 3: Output of MT system.

4.	a.	There is a gun in my bedroom.
	b.	*Hay un revólver en mi alcoba.*
5.	a.	Is there a gun in my bedroom?
	b.	*¿Es allí un revolver en mi alcoba?*
6.	a.	Some of the people over there are Spanish.
	b.	*Alguna de la gente sobre hay Español.*

Source: Based on White (2003, p. 225).

According to the writer (WHITE, 2003, p. 225), in example four shown in the previous table, the input sentence 4(a) “there is” is translated properly by the linking verb in Spanish “*haber*”*,* which is inflected as “*hay*” in the translation generated automatically into Spanish 4(b). However, in example 5 there is an error, which suggests that the translation system recognizes “*haber*” only when the input is exactly in the order “there is*”* or “there are”*.* The suspicion is confirmed by example number 6 because the construction of “there are” in this sentence is different (ibid.).

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

This type of example in the black box perspective is also elaborated by using error typologies often found in a contrasting analysis of what text comes in (input) and what text comes out (output) of the translation systems. Some authors, such as Vilar et. al (2006), use distinct terminologies to analyze phenomena called machine translation "errors" (errors in the sense of machines and system, in opposition to human translation).

In the context of MT error typology studies, the work of Vilar et. al (2006) presents a framework to classify MT errors. This classification expands the work of Llitjós et. al. (2005) and describes five main categories in the following image:

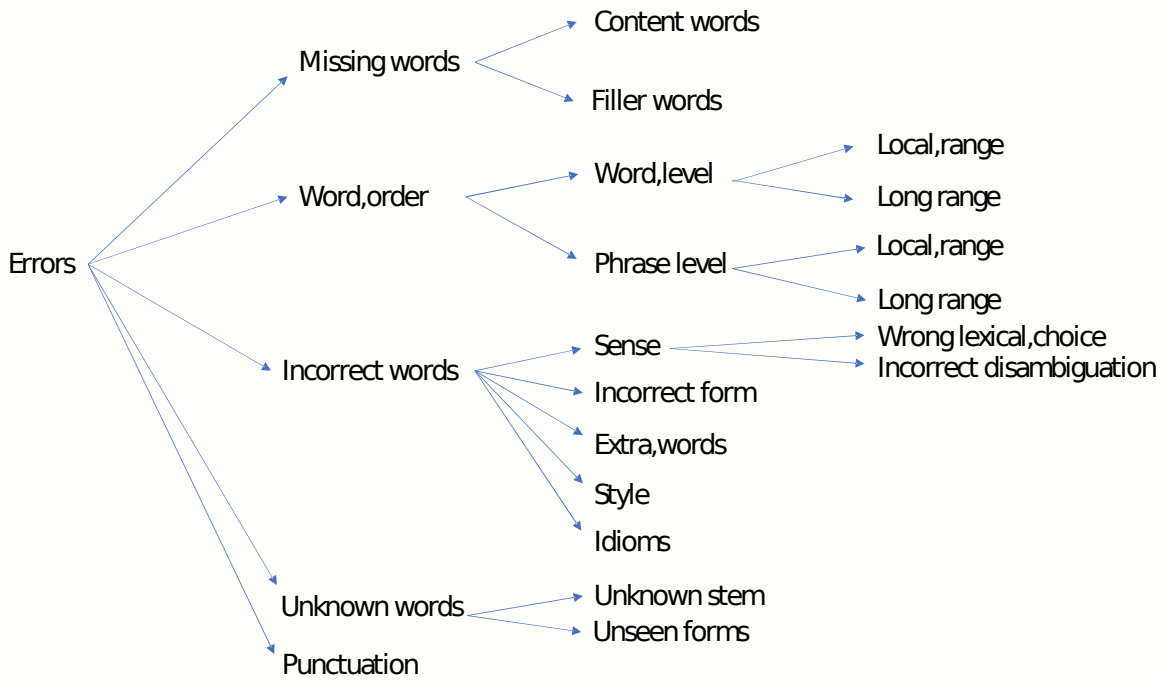

Figure 6: Classification of MT errors.
Source: based on VILAR et al. (2006, p. 699).

As the previous image shows, there are five main MT error categories entitled "missing words", "word order", "incorrect words", "unknown words" and "punctuation".

According to Vilar et. al. (2006, p. 698) the first category represents cases in which a word is missing from sentences generated from MT. Both subcategories, "content words" and "filler words", are respectively needed to express the sense of the sentence and to form the sentence grammatically, though the sense is preserved. The second category is related to the reordering of words and syntactic chunks of words. The difference between both levels relies on the moving of words individually or in chunks of words when generating the sentences. In relation to the local range or long range, the distinction is not made in absolute[8] terms, but relies on the need to reorganize words in a local context (within the same syntactic block) or to move the words to another block (ibid., p. 698).

8 Vilar et. al. (2006, p. 698) highlight that "the classification of the errors of a machine translation system is by no means unambiguous".

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

The next category of the classification scheme (Vilar et. al. 2006, p. 698) describes the “incorrect words” errors, which can be identified when a system does not find an appropriate match for a word. In the first category, an incorrect word interferes in the sense of the sentence, which reveals two subtypes of disambiguation: one, in which the system chooses an incorrect translation (“wrong lexical choice”), and another in which the translation system is not able to disambiguate the proper meaning of word from the source language (“incorrect disambiguation”). The other subcategory of incorrect words is the “incorrect forms”, which occurs when the system does not produce the correct form of a word, though the translation of the basic form is correct. The subsequent subcategory is represented by words added by the generated sentence. The two remaining categories are classified by the “bad” word choices in the automatic translation, though meaning is preserved. Vilar et. al. (2006) don’t consider as completely correct certain stylistic elements, such as the repetition of a word in a close context or idioms which the system don’t recognize, thus generating a translated “normal”[9] text.

The fourth category presented by Vilar et. al. (2006, p. 698) is the “unknown words”, which can be distinguished by truly unknown words or stems, and unseen forms of known stems. The fifth category, “punctuation” is considered by the authors (ibid., p. 698) to be a minor problem for machine translation evaluation.

Such work can be used to describe the main machine translation errors, especially when used within a “black box” evaluation, which humans can classify manually. One relevant work involving such evaluation using Portuguese as one of the languages of a linguistic pair for automatic translation is the TrAva project (Traduz e Avalia[10]). According to Santos et. al. (2004):

> TrAva is thus a system whose goal is to come to grips with some of the intuitively employed criteria of judging translation, by producing a relatively easy framework for cooperatively gathering hundreds of examples classified according to problems of (machine) translations.

According to the project, some descriptions involving manual work with MT error types analysis can be identified and described by Sarmento (2007). His work shows some tools for experimenting with, gathering, and evaluating MT examples. Among other things, Sarmento (2007, pp. 193-203) displays TrAva’s working system schemes and two tables showing the categories for human classification of morphological and lexical problems with automatically translated sentences. As the present study looks at expanding such linguistic classification for manually analyzing new text-image relationship configurations generated by MT output, both tables are displayed here as they are a useful reference for further exploration:

9 Although Vilar et. al. (2006) highlight that error classification is ambiguous and absolute, there are some absolute and semantically loaded terms employed to qualify translation, such as “correct”, “bad” and “normal”.

10 <http://www.linguateca.pt/TrAva/>

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

Table 4: TrAva's Classification of morphological and lexical problems.

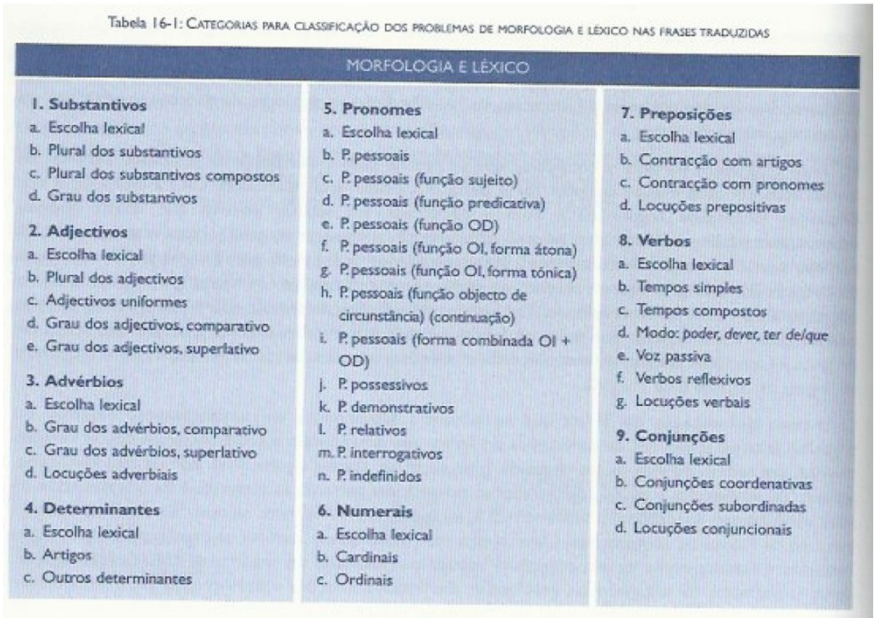

Tabela 16-1: Categorias para classificação dos problemas de morfologia e léxico nas frases traduzidas

MORFOLOGIA E LÉXICO		
1. Substantivos a. Escolha lexical b. Plural dos substantivos c. Plural dos substantivos compostos d. Grau dos substantivos **2. Adjectivos** a. Escolha lexical b. Plural dos adjectivos c. Adjectivos uniformes d. Grau dos adjectivos, comparativo e. Grau dos adjectivos, superlativo **3. Advérbios** a. Escolha lexical b. Grau dos advérbios, comparativo c. Grau dos advérbios, superlativo d. Locuções adverbiais **4. Determinantes** a. Escolha lexical b. Artigos c. Outros determinantes	**5. Pronomes** a. Escolha lexical b. P. pessoais c. P. pessoais (função sujeito) d. P. pessoais (função predicativa) e. P. pessoais (função OD) f. P. pessoais (função OI, forma átona) g. P. pessoais (função OI, forma tónica) h. P. pessoais (função objecto de circunstância) (continuação) i. P. pessoais (forma combinada OI + OD) j. P. possessivos k. P. demonstrativos l. P. relativos m. P. interrogativos n. P. indefinidos **6. Numerais** a. Escolha lexical b. Cardinais c. Ordinais	**7. Preposições** a. Escolha lexical b. Contracção com artigos c. Contracção com pronomes d. Locuções prepositivas **8. Verbos** a. Escolha lexical b. Tempos simples c. Tempos compostos d. Modo: *poder, dever, ter de/que* e. Voz passiva f. Verbos reflexivos g. Locuções verbais **9. Conjunções** a. Escolha lexical b. Conjunções coordenativas c. Conjunções subordinadas d. Locuções conjuncionais

Source: Sarmento (2007, p.202).

Table 5: TrAva's Classification of syntactic problems.

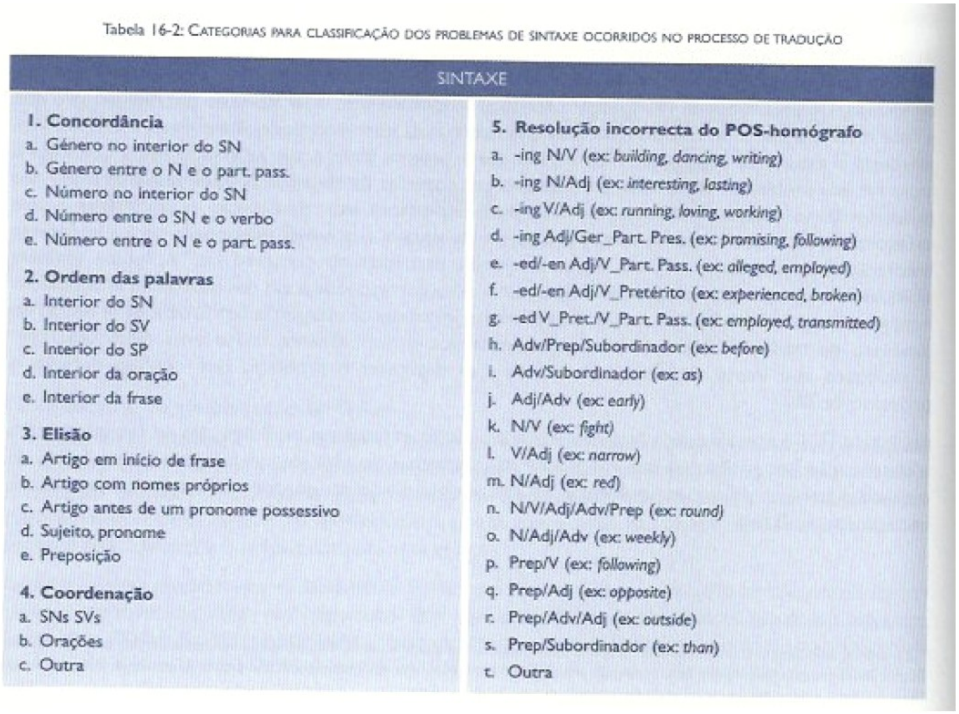

Tabela 16-2: Categorias para classificação dos problemas de sintaxe ocorridos no processo de tradução

SINTAXE	
1. Concordância a. Género no interior do SN b. Género entre o N e o part. pass. c. Número no interior do SN d. Número entre o SN e o verbo e. Número entre o N e o part. pass. **2. Ordem das palavras** a. Interior do SN b. Interior do SV c. Interior do SP d. Interior da oração e. Interior da frase **3. Elisão** a. Artigo em início de frase b. Artigo com nomes próprios c. Artigo antes de um pronome possessivo d. Sujeito, pronome e. Preposição **4. Coordenação** a. SNs SVs b. Orações c. Outra	**5. Resolução incorrecta do POS-homógrafo** a. -ing N/V (ex: *building, dancing, writing*) b. -ing N/Adj (ex: *interesting, lasting*) c. -ing V/Adj (ex: *running, loving, working*) d. -ing Adj/Ger_Part. Pres. (ex: *promising, following*) e. -ed/-en Adj/V_Part. Pass. (ex: *alleged, employed*) f. -ed/-en Adj/V_Pretérito (ex: *experienced, broken*) g. -ed V_Pret./V_Part. Pass. (ex: *employed, transmitted*) h. Adv/Prep/Subordinador (ex: *before*) i. Adv/Subordinador (ex: *as*) j. Adj/Adv (ex: *early*) k. N/V (ex: *fight*) l. V/Adj (ex: *narrow*) m. N/Adj (ex: *red*) n. N/V/Adj/Adv/Prep (ex: *round*) o. N/Adj/Adv (ex: *weekly*) p. Prep/V (ex: *following*) q. Prep/Adj (ex: *opposite*) r. Prep/Adv/Adj (ex: *outside*) s. Prep/Subordinador (ex: *than*) t. Outra

Source: Sarmento (2007, p. 202).

Both tables four and five illustrate the main categories of machine translation problems involving Portuguese. TrAva users employed these tables as criteria to manually evaluate automatic translations, enabling further quantification of the categories. For Sarmento (2007), evaluating translation can be a very arbitrary task, thus the purpose of creating such categories gives concrete selection criteria for evaluating output translation

Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

morphologically, lexically, and syntactically.

Both works explored here are expansions on past investigations on how to catalogue machine translation errors. Although, more than a decade before both works were published, Kameyama et. al. (1991) focused on other aspects of machine translation errors that should be reviewed to serve new purposes such as cataloguing text-image relationship problems caused by MT errors. The authors define the concept called “translation mismatches” identified when “the grammar of one language does not make a distinction required by the grammar of the other language”,

In addition, Kameyama et. al (1991, p. 194) highlight two important consequences of machine translation when there are great mismatches between languages, referring to their contextual information. These consequences are made clear by the authors’ own words:

> Two important consequences for translation follow from the existence of major mismatches between languages. First in translating a source language sentence, mismatches can force one to draw upon information not expressed in the sentence information only inferable from its context at best. Secondly, mismatches may necessitate making information explicit which is only implicit in the source sentence or its context.

The joint study of translation mismatches and classification schemes that focus on manually classifying MT error types presented in this section have is a promising research relation to engage efforts on the phenomenon of intersemiotic mismatches automatically generated by machine translation in multimodal documents.

With some relevant points explored on multimodality and the evaluation of machine translation, now it is time to turn our focus to the potential they have for exploring text-image relationships generated by MT output.

5 What could both approaches do together?

The previous sections explored some potential aspects of multimodality and the evaluation of machine translation considering the research problem described and contextualized in the second section. Such an investigative proposal may be visually represented as follows:

http://periodicos.letras.ufmg.br/index.php/textolivre
Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

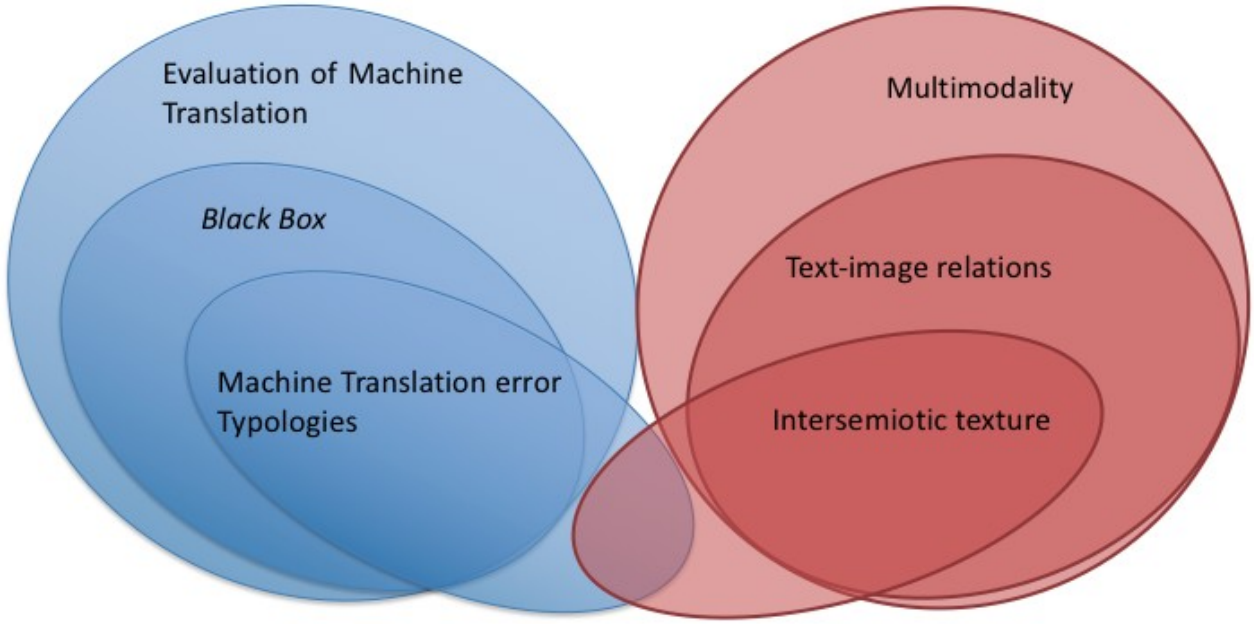

Figure 7: Proposal of interdisciplinary interface.
Source: PIRES (2017).

As one can notice, the previous image demonstrates the interdisciplinary interface proposed in this paper. The subareas of each theoretical background (i.e. evaluation of machine translation and multimodality) meet at the intersection where the problem to be investigated is located; that is where text-image relationships emerge from multimodal documents (e.g. webpages, illustrated manuals and infographics) translated automatically (PIRES, 2017).

But what can both approaches do together as an effort to accommodate the research problem described in this study? There are some aspects in Liu and O'Halloran (2009) on the concept of "intersemiotic texture" that contribute by offering categories based on a systemic-functional social semiotics that focus on the linguistic part within the verbal mode, from which can be gathered either its grammar or discourse to compare to part of the visual mode.

It is exactly within that verbal mode that categories aimed at evaluating machine translation within a "black box" method that (to evaluate its quality) contribute to informing which part of the verbal mode has changed. Thus, a possibly new text-image relationship configuration emerged by using machine translation could be linguistically detected by using MT error typology.

In this sense, not only could Liu and O'Halloran's (2009) intersemiotic texture categories be expanded, but also Sarmento's (2007) MT error classification for human evaluation campaigns to classify new configurations of text-image relationships in multimodal documents translated automatically. It should be highlighted though, that the purposes of Sarmento's (2007) classification and the present study differ in the sense that the latter aims at using such resources for exploratory reasons, rather than using the evaluation of machine translation results to improve MT precision.

Therefore, from the concept of "translation mismatches" (KAMEYAMA et. al., 1991)

Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

the present study attempts to propose the addition of the intersemiotic element to manual MT error classification so as to identify new text-image relationships emerging from machine translation output of multimodal documents, namely “intersemiotic mismatches”.

Such a proposal has a substantial potential to describe such phenomena by providing mutual feedback from both areas, with aims at, perhaps, creating a new interdisciplinary one, such as with the communicative view, involving social semiotics and language with the intersemiotic gap left by automated translation processes of multimodal documents.

Pires (2017) has shown some preliminary findings on such matters, but more research is expected to replicate and investigate the possibility of a pattern and how it is formed. For that reason, future works including a variety of genres, machine translation, and annotation software that help the analyst to work a large number of text-image relationships might present a valuable contribution to examine and describe such problems with more clarity.

The present study in no way assumes to be able to describe all possible elements that multimodality, evaluation of machine translation, and the combination of both could produce together. Rather, it briefly proposed some potentialities within the boundaries of the research problem delimited in this work.

Acknowledgment:

This research was partially supported by Brazilian Funding Agency CAPES, granting a research visit to professor John Bateman’s research group at the University of Bremen, Germany in 2015. I would like to show my gratitude to John for kindly sharing his wisdom during this period, and thank everyone who made part of his research group, kindly supporting my studies.

References

BATEMAN, J. A. *Multimodality and genre:* a foundation for the systematic analysis of multimodal documents*.* New York: Palgrave MacMillan, 2008.

BATEMAN, J. A. *Text and image:* a critical introduction to the visual/verbal divide*.* New York: Routledge, 2014.

BAZERMAN, C. Systems of genres and the enactment of social intentions. In: *Genre and the new rhetoric.* London: Taylor and Francis, 1994. p. 79–104.

DE BEAUGRANDE, R. A.; DRESSLER, W. U. *Einfurung in die Textinguistik*. Tubingen: Niemeyer, 1981.

FAIRCLOUGH, N. *Discourse and social change*. Cambridge: Polity press, 1992.

HALLIDAY, M. A. K. *Language as social semiotic:* towards a general socialingsocial theory. Columbia: Hornbeam Press, 1975.

Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

HALLIDAY, M. A. K. *Language as social semiotic:* the social interpretation of language and meaning. London: Arnold, 1978.

HALLIDAY, M. A. K. *Introduction to functional grammar*. Londres: Arnold, 1985.

HALLIDAY, M. A. K. *Introduction to functional grammar*. 2. ed. Londres: Arnold, 1994.

HALLIDAY, M. A. K.; HASAN, R. *Cohesion in English*. London: Longman, 1976.

HALLIDAY, M. A. K.; MATTHIESSEN, C. *An introduction to functional grammar*. [s.l.] Arnold, 2004.

IEDEMA, R. Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual Communication*, v. 2, n. 1, p. 29–57, 2003.

JEWITT, C. *The Routledge handbook of multimodal analysis*. London: Routledge, 2009.

JONES, J. 2006. *Multiliteracies for academic purposes:* a metafunctional exploration of intersemiosis and multimodality in university textbook and computer-based learning resources in science. Doctoral thesis, University of Sidney , Sidney, Australia, 2006.

KAMEYAMA, M.; OCHITANI, R.; PETERS, S. Resolving translation mismatches with information flow. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS ACL91, 1991. *Proceedings of the Annual Meeting of the Association for Computational Linguistics ACL91*. 1991.

KALTENBACHER, M. Perspectives on multimodality: From the early beginnings to the state of the art. *Information Design Journal*, v. 12, n. 3, p. 190–207, 2004.

KRESS, G.; VAN LEEUWEN, T. *Reading images:* the grammar of visual design. 1st. ed. London: Routledge, 1996.

KRESS, G.; VAN LEEUWEN, T. *Multimodal discourse:* the modes and media of contemporary communication. London: Edward Arnold, 2001.

KRESS, G.; VAN LEEUWEN, T. *Reading images:* the grammar of visual design. 2nd. ed. London: Routledge, 2006.

LEMKE, J. Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text. In: *Reading science*. Londres: Routledge, 1998. p. 87–113.

LEMKE, J. Typology, topology, topography: genre semantics. *MS. University of Michigan*, 1999. Disponível em: <http://academic.brooklyn.cuny.edu/education/jlemke/papers/Genre-topology-revised.htm>. Acesso em: 7 abr. 2018.

LIU, Y.; O'HALLORAN, K. L. Intersemiotic texture: analyzing cohesive devices between language and images. *Social Semiotics*, v. 19, n. 4, p. 367–388, 2009.

Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018 – ISSN 1983-3652
DOI: 10.17851/1983-3652.11.1.82-102

LLITJ´OS, A. F.; CARBONELL, J. G.; LAVIE, A. A framework for interactive and automatic refinement of transfer-based machine translation. In: ANNUAL CONFERENCE OF THE EUROPEAN ASSOCIATION FOR MACHINE TRANSLATION (EAMT). 2005, Budapeste. *Proc. of the 10th Annual Conf. of the European Association for Machine Translation (EAMT)*, 2005.

MARTIN, J. R. English text: system and structure. [s.l.] John Benjamins Pub. Co, 1992.
MARTINEC, R.; SALWAY, A. A system for image-text relations in new (and old) media. *Visual Communication*, v. 4, n. 3, p. 337–371, 2005.

O'HALLORAN, K. L. *Mathematical discourse:* language, symbolism and visual images. London: Continuum, 2005.

O'TOOLE, M. *The language of displayed art.* London: Leicester University Press, 1994.

PIRES, T. B.; DUQUE, C. G. Sistemas de gerenciamento de tradução: uma proposta de análise multimodal. *DataGramaZero,* v. 16, n. 3, 2015.

PIRES, T. B. *Ampliando olhares sobre a tradução automática online:* um estudo exploratório de categorias de erros de máquina de tradução gerados em documentos multimodais. 2017. 169f. Doctoral thesis – Programa de Pós-Graduação em Ciência da Informação, Unb, Brasília, 2017.

ROYCE, T. Synergy on the Page: Exploring intersemiotic complementarity in page-based multimodal text. *JASFL Occasional papers*, v. 1, n. 1, p. 25–49, 1998.

ROYCE, T. New directions in the analysis of multimodal discourse. In: *New directions in the analysis of multimodal discourse*. New York: Routledge, 2007. p. 63–109.

SANTOS, D.; MAIA, B.; SARMENTO, L. Gathering empirical data to evaluate MT from English to Portuguese. In: WORKSHOP ON THE AMAZING UTILITY OF PARALLEL AND COMPARABLE CORPORA. 2004, Lisboa, Portugal. *Proceedings of LREC,* Lisboa, 2004.

SARMENTO, L. Ferramentas para experimentação, recolha e avaliação de exemplos de tradução automática. In: *Avaliação conjunta:* um novo paradigma no processamento computacional da língua portuguesa. Lisboa, Portugal: IST press, 2007. p. 193–203.

VILAR, D. et al. Error analysis of statistical machine translation output. In: LREC, 2006, Genoa. *Proceedings of LREC*. Genoa, 2006.

WHITE, J. S. How to evaluate machine translation. In: *Computers and Translation:* A translator's guide. Amsterdam/Philadelphia: John Benjamins Publishing, v. 35, p. 211–244, 2003.

Recebido em dia 08 de abril de 2018.
Aprovado em dia 03 de maio de 2018.