Challenges and Difficulties in the Development of Dicionário Olímpico (2016)

Rove Chishman, Aline Nardes dos Santos, Bruna da Silva, Larissa Brangel

Unisinos University, São Leopoldo, Brazil
E-mail: rove@unisinos.br, aline.nardes@gmail.com, broonamoraes@gmail.com, larissabrangel@gmail.com

Abstract

This paper discusses some theoretical and practical implications arising from the development of the Dicionário Olímpico (2016), created by the SemanTec (Semantics & Technology) research group. The Dicionário Olímpico (available at http://www.dicionarioolimpico.com.br/) is a bilingual lexicographic resource (Portuguese-English) which describes the lexicon of 40 Olympic sports. The dictionary is based on the theoretical-methodological framework of Frame Semantics, developed by Charles J. Fillmore. The paper brings some background to the Dicionário Olímpico's methodological approach. In addition, it describes the lexicographical structure of the resource and the way frame-semantic features were incorporated and adapted in this context. Finally, it explores two kinds of challenges faced by the project: the identification and description of semantic frames, and the design of a template for frame definitions. These stages of development have included some adaptations of frame-semantic concepts with the purpose of building a user-friendly, frame-based dictionary. Such challenges have enriched the lexicographic work and impacted subsequent projects that are yet to be developed by the authors.

Keywords: Frame Semantics; Frame-based dictionary; Dicionário Olímpico.

1. Introduction

The contributions of Frame Semantics (Fillmore, 1982, 1985) to lexicography have been widely addressed since Fillmore's first research works within the context of FrameNet Berkeley, the first frame-based lexicographical database ever published (https://framenet.icsi.berkeley.edu/). For example, Atkins, Rundell and Sato (2003) and Atkins, Fillmore and Johnson (2003) approached the contributions of FrameNet to practical lexicography, especially in the process of managing and manipulating corpus data to extract lexicographically relevant information. In this regard, Fillmore and Atkins (1992:75) explored the idea of building an online frame-based lexicographical resource: "In such a dictionary [...], individual word senses, relationships among the senses of polysemous words, and relationships between (senses of) semantically related words will be linked with the cognitive structures (or 'frames'), knowledge of which is presupposed for the concepts encoded by the words."

More recently, advances towards a richer convergence between Frame Semantics and dictionary writing have increased. Specifically, we highlight the works by Ostermann (2012, 2016) concerning Cognitive Lexicography and the improvement of dictionary sections by the inclusion of information based on cognitive theories. In this sense,

practical lexicography imposes many challenges when it comes to articulating cognitive-linguistic theories such as Frame Semantics with dictionary-making processes, since "The craft of lexicography demands not only the ability to collect data, [...] we need to set out these facts in an intelligible and orderly way." (Atkins, 2002: 171).

This paper aims at discussing some of these challenges within the context of development of the Dicionário Olímpico (DO) (http://www.dicionarioolimpico.com.br/), a bilingual lexicographic resource (Portuguese-English) which describes the lexicon of 40 Olympic sports. The dictionary is based on the theoretical-methodological framework of Frame Semantics (Fillmore, 1982, 1985). More specifically, the paper approaches some of the challenges faced by the developers during the process of compilation of the Dicionário Olímpico, considering that such challenges have enriched the lexicographic work and impacted on subsequent projects that are yet to be developed by the authors. The rest of the paper is structured as follows. Section 2 provides some background to the development of the DO, including its methodological approach. Section 3 describes the lexicographical structure of the dictionary and the way frame-semantic features were incorporated and adapted in this context. Section 4 focuses on two kinds of challenges faced by the project: identification and description of semantic frames (section 4.1), and the ongoing design of a template for frame glosses (section 4.2).

2. Background to the Dicionário Olímpico

The Dicionário Olímpico is a Brazilian bilingual dictionary of Olympic sports developed within the context of the 2016 Olympic Games. It is the result of a broader academic project whose purpose was to study the potential convergence between Frame Semantics and lexicography for the purpose of describing the lexicon of sports. Two years earlier, the research group responsible for building this resource had already launched a frame-based football dictionary called Dicionário Field (http://dicionariofield.com.br), a trilingual resource (in English, Spanish, and Portuguese) structured by semantic frames. During this first lexicographical project, among other results, the group explored the relevance of Frame Semantics for lexicographical practice, not only in terms of enhancing the process of collecting lexicographically relevant information (Chishman et al., 2015), but also with regard to making a dictionary more contextualized by duplicating its macrostructure and enabling users to look up words, frames and different evokers of the same scenario (Santos & Chishman, 2015).

Although some frame-semantic assumptions are adapted in these projects (see Section 3.1 for more details), it is important to approach the theory's core concepts that underlie the building of the Dicionário Field and the Dicionário Olímpico, enriching their content and access structures. According to Fillmore and Baker (2010: 237), Frame Semantics assumes that "[...] the meaning dimension is expressed in terms of the cognitive structures (frames) that shape speakers' understanding of linguistic expressions." For example, in football, a word such as *assist* can only be understood if a speaker recognizes the cognitive structure it evokes, which is constituted of encyclopaedic and sociocultural information:

in football, a player *assists* a scorer of a goal when he passes him the ball. Therefore, to *assist* means to supply a specific (and decisive) kind of pass in football – hence this word evokes the Pass frame. As Fillmore (1985: 229) states, "Frame semantics allows the possibility that speakers can have full knowledge of the meaning of a given word in a domain [...]". In other words, understanding a word (or, in frame-semantic terms, understanding a lexical unit) implies recognizing the frame it evokes.

The challenge of describing the language of sports through semantic frames became bigger with the development of the Dicionário Olímpico. Firstly, while Field is a football dictionary, DO describes 40 Olympic sports. Secondly, the corpus compilation imposed other difficulties: to build a corpora for the basis of Field's lexicographical work, the editors selected match reports from football websites, which is a pervasive text genre both in Brazilian Portuguese and in English (more specifically, those on British websites). However, in the context of the Dicionário Olímpico, only a few Olympic sports, such as volleyball and basketball, are as popular as football in Brazil; thus match reports could not be used as the main sources to build all corpora. In case of less popular games, sometimes the only reliable written documents available concerned the rules of these sports.

Therefore, in order to broaden the range of text genres for corpus compilation purposes, the following procedures were adopted: transcription of match videos available online; compilation of documents such as sports rules and other official materials; and a qualitative study of sports-related videos and other multimodal materials whose content was not processable by a corpus tool, nor worth transcribing – since it is a very time-consuming task. Indeed, these multimodal sources provided supporting information and were used as a reference material for comparing and complementing the study corpora.

The Dicionário Olímpico's corpora were processed and managed through Sketch Engine. This tool is renowned for its relevance for dictionary writing, especially due to the word sketches it provides, which "combine information of two types: grammatical relations in the corpus, and statistically significant frequencies of co-occurrence" (Atkins et al., 2003: 336). As described by Chishman et al (2017), after planning the macro- and microstructure of the dictionary, the development of the Dicionário Olímpico included the following stages: (i) study of sports and systematization of their main characteristics; (ii) corpus design and compilation, including documents such as sports rules and match reports, if available; (iii) gathering of multimodal supporting material, especially in case of little-known sports; (iv) creation of conceptual maps regarding the respective domains, which were based on the previously collected written corpora and multimodal sources; (v) description of semantic frames, based on the previously designed and discussed conceptual maps; (vi) corpus extraction of possible lexical frame evokers and their equivalents; (vii) writing and collective revision of the entries by the editors, with the assistance of sports experts (for example, coaches and former players); (vii) building of the entries on the dictionary website database. All these stages have brought many challenges that have been, or are yet to be, discussed.

3. The Dicionário Olímpico: lexicographic structure

As we saw earlier, since the Dicionário Olímpico (DO) was developed from the theoretical-methodological framework of Frames Semantics, many aspects of the lexicographic structure of this tool were based on FrameNet's lexicographic structure. However, there is only a slight degree of similarity between these two tools, since the target audiences also differ.

At this point, it is relevant to mention that the target audience consists of people who relate directly to the Olympic modalities, such as students, athletes and other sports professionals; and also includes users whose relationship with Olympic sports is indirect, such as translators and people interested in this topic. Above all, the DO audience includes people who do not necessarily have any extensive knowledge of linguistics' or lexicography's theoretical concepts.

With this in mind, in this section, we describe the lexicographic structure of the DO: how to access data and the levels of the dictionary. In addition, we discuss how the notion of frame has been incorporated into the project, emphasizing the centrality of the intended audience in the process of definitions regarding the content and form of the dictionary.

3.1 Access to data

Considering that the DO is composed of 40 dictionaries, each one corresponding to one of the sports that comprise the framework of the Summer Olympics, the resource's homepage enables users to select a specific Olympic modality (from the respective icons) or the search for a word, scenario, or modality (from the search box), as shown in Figure 1.

At this point, it is necessary to approach the first adaptation that was necessary in the development of the DO. In the context of FrameNet, the terminological concepts 'frame' and 'lexical unit' are used. This is due to the fact that the target audience comprises predominantly linguistics researchers, teachers, and students, i.e., people who are familiar with these theoretical concepts.

On the other hand, the potential audience of the Dicionário Olímpico is composed of non-specialists. For this target audience, the use of theoretical concepts could lead to a communication failure. With that in mind, the SemanTec research group adopted words that sound more familiar to the user. Thus, the word 'frame' was replaced by 'scenario', and the expression 'lexical unit' was replaced by 'word' in the structure of the dictionary.
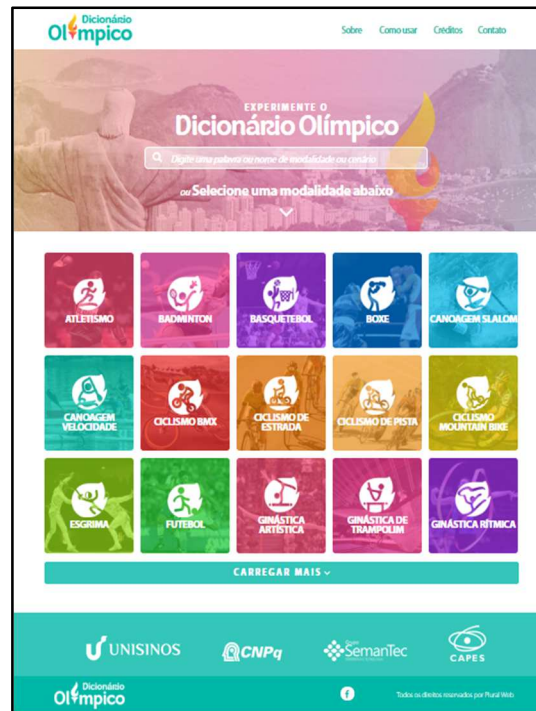
Figure 1: Dicionário Olímpico's homepage

3.2 Access levels of the Dicionário Olímpico

When selecting one of the forms of access, users are directed to one of the three levels of the DO: the modality level, the scenario level, or the word level. Each of them is presented in the following subsections.

3.2.1 First level: Olympic modality

When selecting one of the sports on the homepage, users are directed to a page containing this set of information: gloss (supergloss), conceptual map, scenario list, word list, trivia section, related sports, and image, as shown in Figure 2.

The most significant differences between the Dicionário Olímpico and FrameNet are at this level. While FrameNet describes general frames, the Olympic Dictionary describes the frames of Olympic sports, which are called, in this context, superframes. Thus, each frame of the Dicionário Olímpico corresponds to an Olympic modality, and not to the Olympic domain as a whole. In contrast, FrameNet does not group frames by domains.

For this reason, this level presents elements that do not exist in FrameNet, such as conceptual maps for each Olympic modality and the trivia section, which are a result of decisions made during the Dicionário Olímpico development process. The reasons for these decisions are explained in the next sections.
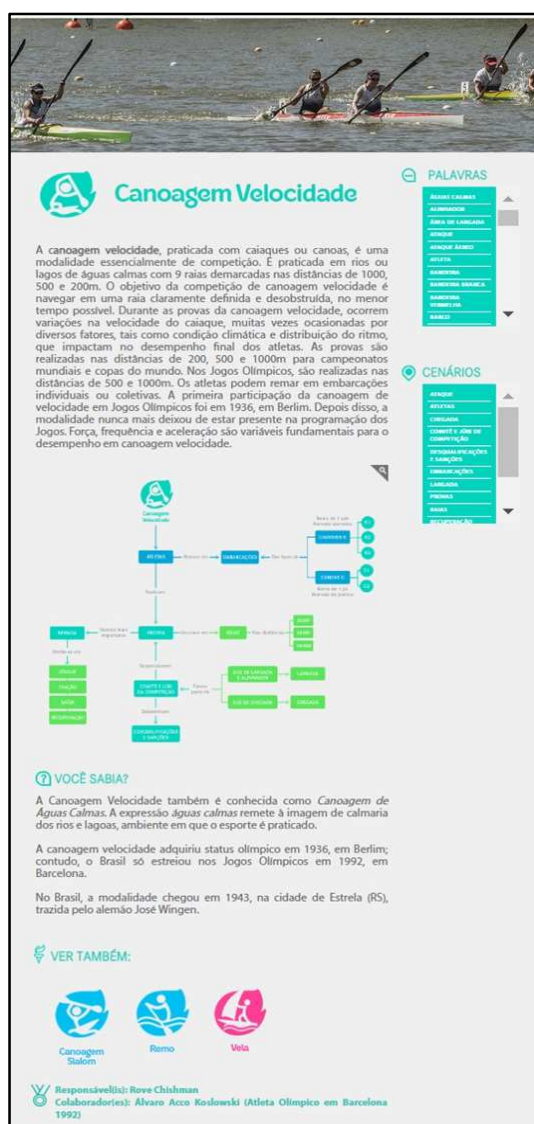
Figure 2: Level of the Olympic modality or superframe

3.2.2 Second level: scenario

In terms of content, the level of the scenario resembles the modality level. The elements that constitute it are: gloss, list of words, related frames, image, and conceptual map, as shown in Figure 3.

Elements that were based on FrameNet's structure, such as gloss, word list, and relations between scenarios, have undergone some modifications. Regarding the relations between scenarios, it is worth mentioning that initially the editors intended to use the set of frame relations created by FrameNet: inheritance, perspective, use, subframe and precedence. However, on submitting the dictionary content to the experts' inspection, the research group received negative feedback. According to these professionals, these relations were obscure; they were not user-friendly. For this reason,

FrameNet relations were not used, and those responsible for each Olympic sport were in charge of identifying the types of relations that could be established between the frames, based on the study of each discipline.
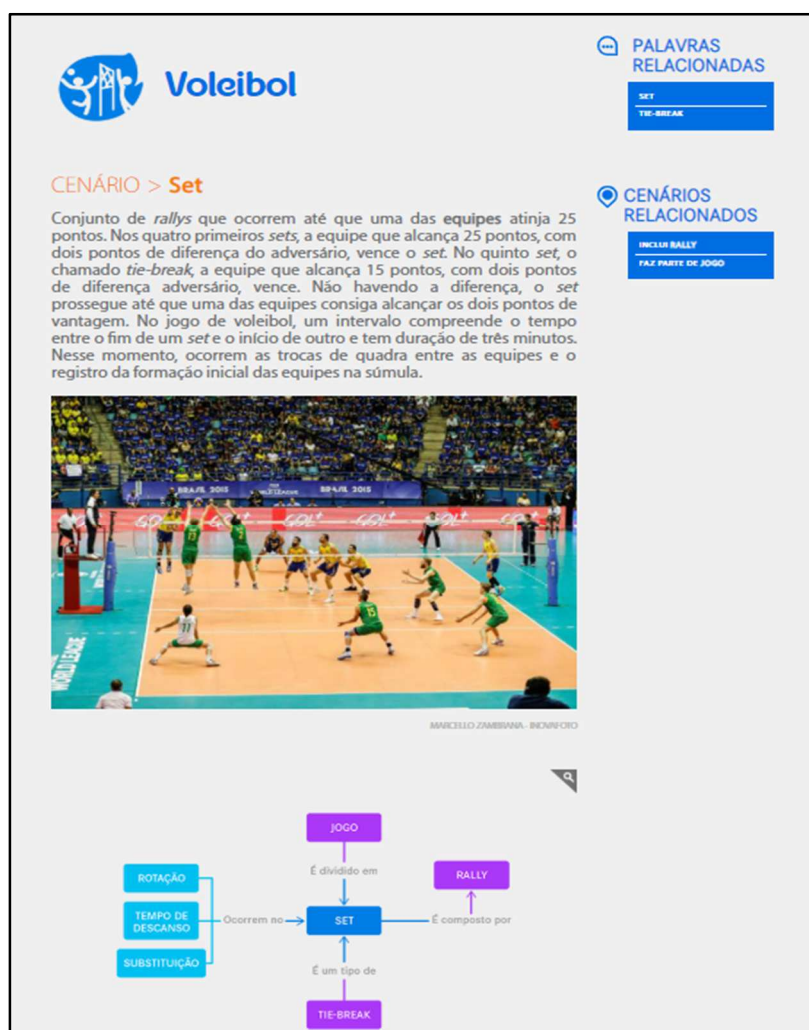

Figure 3: Scenario level.

The following figure presents the relations for the basketball frame called Basket: basket depends on Shot; generates Throw-in; uses Team; and uses Court. Other types of relations used in this context were 'part of', such as in the badminton frame Equipment (Equipment is part of Court); 'to control', as in the beach volleyball frame Refereeing (Refereeing controls the Match); and 'to execute', as in the tennis frame Tennis Players (Tennis Players execute Shot).

CENÁRIO > Cesta

O principal objetivo de uma equipe em um jogo de basquete é obter pontos ao fazer com que a **bola** passe por dentro do **aro** localizado na quadra de ataque. A conversão de um lance livre vale 1 ponto. Os arremessos convertidos durante as jogadas normais de jogo – as chamadas cestas de campo – podem valer 2 pontos, quando originadas de arremessos convertidos de dentro da linha de três pontos, ou 3 pontos, quando provenientes de arremessos realizados de fora da linha de três pontos.

CENÁRIOS RELACIONADOS

DEPENDE DE **ARREMESSO**
GERA **REPOSIÇÃO DE BOLA**
USA **EQUIPE**
USA **QUADRA**

Figure 4: Relations between scenarios

Furthermore, the glosses of the Dicionário Olímpico, an element that is discussed in more detail in the next section, do not follow the structure of FrameNet's standard glosses, which are built through the following steps: (i) characterizing the frame; and (ii) describing and naming frame elements (Fillmore & Baker, 2009).

Figure 5: FrameNet gloss model

In Dicionário Olímpico, it was considered that these elements would not receive the prominence they have in FrameNet. Instead, glosses – both glosses (scenarios) and superglosses (Olympic modalities) – feature prominent words that are not necessarily frame elements, but can be viewed as keywords that are necessary to understand the respective frame.

CENÁRIO > Resultado

O resultado de um **jogo** de badminton pode ser parcial ou final. O **resultado parcial** envolve o número de **pontos** marcados pelos jogadores durante um *set*. O **resultado final** envolve o número de *sets* vencidos por cada jogador ou equipe ao final do jogo. Para vencer um set, um jogador ou dupla precisa conquistar marcar 21 pontos, com dois pontos de diferença do adversário. Para vencer um jogo, os jogadores precisam conquistar o **melhor resultado de 3** *sets*.

Figure 6: Gloss model of the Dicionário Olímpico

Figure 7: Dicionário Olímpico supergloss model

The structure of the modality and the scenario levels resemble each other, according to their nature. In the context of the DO, modalities are considered more comprehensive frames (superframes), and for that reason they should be described in a similar way to how are described.

An element that integrates the levels of the Olympic modality and the scenario is the conceptual map. Initially used only as a methodological strategy for the organization of information about modalities, the conceptual maps were later included in the access structure because they include, albeit implicitly, some notions underlying Frame Semantics.

The task of connecting frames and frame elements refers to the notion of frames as sets of related concepts, in such a way that to understand one of them it is necessary to understand the system as a whole (Fillmore, 1982). Thus, by locating a frame or frame element on the conceptual map, users identify the role that such unit plays within the system. In addition, the way these relations between concepts are presented refers to FrameNet's frame-to-frame relations. From this information, users identify the ways in which, for example, one frame contributes to a preceding one or how a frame integrates a larger one (subframe).

Finally, the inclusion of images (modality level and scenario level), the trivia section and "see also" (both at the modality level) aim at meeting the encyclopaedic character of the dictionary. Images, for example, play a role as frame evokers. The "see also" section, in turn, highlights the similarities between sports whose structures share the same bases (for example, rhythmic and artistic gymnastics).
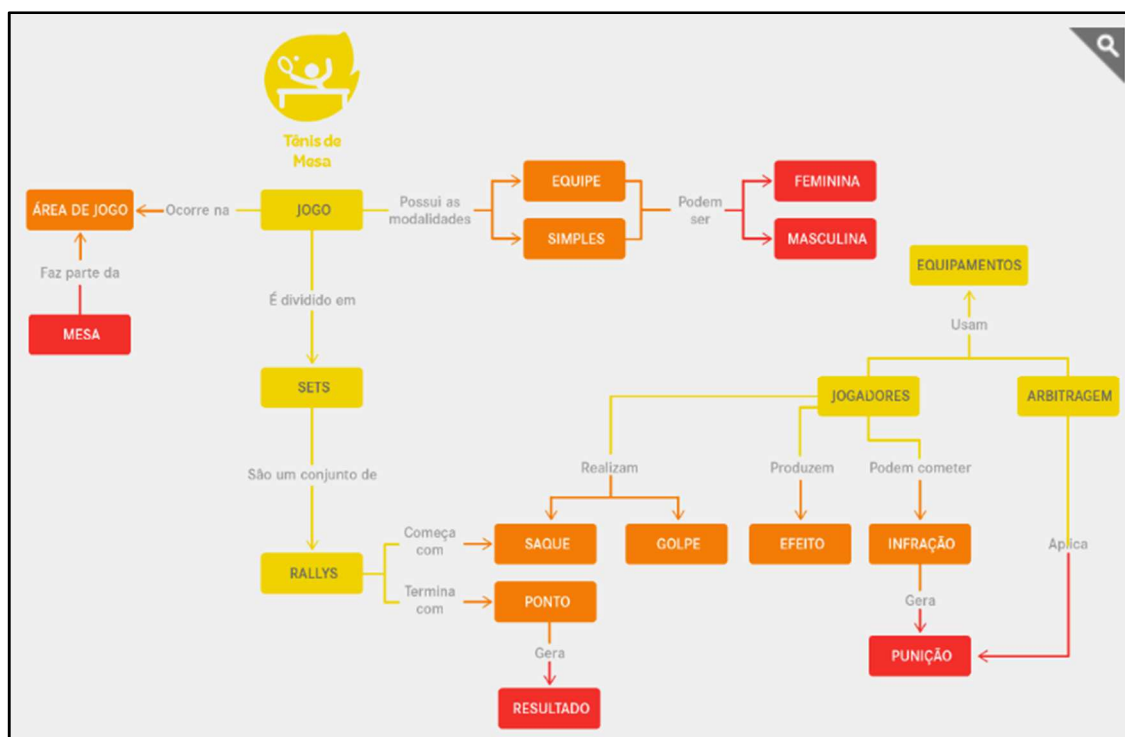
Figure 8: Conceptual map of table tennis

3.2.3 Third level: word

The third and last level of the DO presents information related to the words of the Olympic modalities. From this level, users have access to the grammatical classification of the word, the scenario which the word searched evokes, the English equivalent, an example and a list of other words that integrate the correspondent scenario. Notes are presented in some cases, for the purpose of providing more specific information about a word. In addition, variants are presented when the same phenomenon can be named in two or more different ways.

Figure 9: Level of word 1

Figure 10: Level of word 2

In comparison to FrameNet, the DO's lexical unit entries do not include features such as semantic type, frame elements and their syntactic realizations, and valence patterns. The editors considered that such information could represent an overly theoretical level, considering the intended audience for the dictionary. Other kinds of information that was suppressed concerned the definition of lexical units. However, the notes on the DO have a similar function to FrameNet's definitions.

Variants and translation equivalents were also proposed. Regarding variants, it is worth noting that their use was quite broad in the dictionary. This was due both to the regional differences in Brazil and to the fact that Olympic sports that are not widespread in the country present many words in English which have not yet been fully adopted in Brazilian Portuguese.

In this section, we presented the lexicographic structure of DO, showing the similarities and differences that this tool presents in comparison to FrameNet. We intended to highlight the reasons that led to adaptations of some of the FrameNet's features and to the inclusion of new elements in the dictionary. In the next section, we address some of the key challenges faced in the process of developing the DO and discuss how we dealt with such difficulties.

4. Challenges in the development of the Dicionário Olímpico

As demonstrated in the previous sections, the lexicographical structure of the DO provides the user with all the modalities of the Olympic sports in the form of superframes. Inside each superframe (that can be accessed by a hyperlink), the user can find a set of information about the sports, such as images, conceptual maps, lists of words, lists of frames and glosses. This section presents a brief overview about the challenges and difficulties faced by the SemanTec group during the description of frames: identification and description of semantic frames (section 4.1), and writing of

the glosses (section 4.2). Moreover, these sections discuss how such difficulties have been circumvented in the compilation of Dicionário Paralímpico, a dictionary of the Paralympic sports that is currently under development.

4.1 Identification and description of semantic frames

In a frame-based dictionary, all structural elements are somehow subordinate to the set of frames described. Therefore, among the tasks of compiling a dictionary of this nature, the frame definition step occupies a central position, since it is the stage from which the dictionary begins to be constructed.

As we saw in the previous section, the DO compilation process comprised a series of adaptations of FrameNet's lexicographic model. In this regard, one of the stages that was not based on FrameNet's methodology was the step of identification of frames.

Regarding FrameNet, Fillmore and Baker (2009: 320) state that "The method of inquiry is to find groups of words whose frame structures can be described together, by virtue of their shared common schematic backgrounds." However, the method of identifying frames used by FrameNet compilers is not explicit. In describing the process of lexical analysis of the platform, for example, the authors begin the process of frame identification with a step related to the characterization of the frame.

It is important to highlight that not even the frames already described by FrameNet could be used as a starting point to describe Olympic sports' scenarios, since FrameNet does not describe frames of more specific domains. In addition, establishing the frameset of a general language and describing frames from a sports domain are not equivalent activities. With this in mind, the identification of the Olympic frames started from scratch and can therefore be considered one of the most challenging tasks performed during the development of the DO.

Therefore, the SemanTec research group outlined a methodology for identifying the frames based on the conceptual mapping of the Olympic modalities. A similar procedure was used in the development of the Field dictionary. However, it was in the context of the Dicionário Olímpico that the use of this methodology acquired more definite contours.

This procedure was constituted of two steps: elaboration of the general map and design of the map of the frames. In the first stage, the editors in charge of the description of each Olympic modality elaborated a more comprehensive conceptual map, describing the sports with a high level of detail. In order to do so, the group studied the support materials, mentioned in the beginning of this paper, from which detailed information of the sports, including terms, expressions and specific concepts, were extracted. At this stage, conceptual maps eventually incorporated the organizational structure of sports manuals, since many titles and sections of these materials were converted into central map nodes.

In the second stage, from the more general map, it was possible to design a conceptual map containing only the Olympic sport's frames. In this process of refinement, the objective was to build the final conceptual map of each Olympic modality and to establish a definitive list of frames. The main methodological procedure for this step was the systematization of the list of lexical units, in order to divide them into groups of words that together evoked the frames of each sport. As a final step, the material was sent to an expert.

In view of the innovation represented by the use of conceptual maps in a frame-based dictionary, and considering the lack of methodological support for the elaboration of these maps, the strategies described above represent a first step towards dealing with challenges of this nature. Currently, in the process of compiling the Dicionário Paralímpico, the group has been discussing new forms of frame identification, in order to improve this method and to evaluate the most efficient methodological procedures.

4.2 The glosses

First of all, it is necessary to define what we understand by gloss in the context of the DO. In semasiological dictionaries, a gloss is usually regarded as "a paraphrase or synonym used within a dictionary entry to provide an explanation of the sense of a word or phrase related to the headword" (Hartmann & James, 2002, s.v. gloss). This is not, however, an applicable definition to the glosses of the Dicionário Olímpico, which are, in fact, brief texts located in specific sections of the dictionary with the purpose of providing the user with information about the Olympic sports. Once the glosses of the DO comprise information classified as "encyclopaedic", they are more closely related to an encyclopaedic definition: "a definition which reflects encyclopaedic knowledge (about facts) rather than linguistic knowledge (about words)" (Hartmann & James, 2002, s.v. encyclopaedic definition). The term "definition", however, is still often related to the brief explanations found in the entries of semasiologic monolingual dictionaries, and this is the reason why we gave a proper nomenclature to the textual information about the sports in DO: gloss.

The glosses of DO are located in two specific parts of the dictionary: in modality entries and in frame entries. In the modality entries, the glosses provide the user with a set of information about a specific Olympic sport, helping them to know the main facts and features about the sport. This part of the dictionary is called superframe (as referred in Section 3.2.2), and this kind of gloss is called supergloss. In the frame entries, glosses intend to describe the frames of each sport, helping users to understand some specifics of each Olympic modality, such as the equipment used and the rules of the games. The figures below present the supergloss of artistic gymnastics and the gloss of one of its frames:

Ginástica Artística

A ginástica artística é um esporte de precisão formal no qual os ginastas devem apresentar uma **série** composta por **elementos** acrobáticos e ginásticos em um dos **aparelhos** que integram a competição. Na competição feminina, as ginastas se apresentam em quatro aparelhos: as **barras assimétricas**, o **salto sobre a mesa**, o **solo** e a **trave de equilíbrio**. Na competição masculina, os ginastas também competem no **salto sobre a mesa** e no **solo**, além de executarem séries nas **argolas**, na **barra fixa**, nas **barras paralelas** e no **cavalo com alças**. Um **júri** composto por 8 árbitros avalia o ginasta com relação à **dificuldade** da série (com base nos valores dos elementos que a compõem, estabelecidos pelo Código de Pontuação) e à **execução** da série (conforme a qualidade e precisão técnica dos movimentos realizados pelo ginasta). Ao apresentar a série, os ginastas podem cometer falhas, o que acarreta em **dedução** na nota, ou realizar movimentos combinados ou de grande dificuldade, o que resulta em **bonificação**. Esses itens determinam a **pontuação** da série do ginasta. Geralmente, as séries apresentam uma **entrada**, forma de iniciar a apresentação e o contato com o aparelho, a execução dos elementos que compõem a série e uma **saída**, finalização da série e término do contato com o aparelho. Diferentes tipos de **elementos** são utilizados nas acrobacias realizadas pelos ginastas nos aparelhos. **Saltos mortais, piruetas, saltos de dança**, e **apoios** são alguns dos elementos que, realizados em **sequência**, compõem as séries. Presente na competição desde a primeira edição dos Jogos Modernos, em Atenas 1896, as provas acontecem nas categorias individual geral, por equipes e por aparelho. Nas finais, competem apenas os 8 ginastas ou equipes que obtiverem as melhores notas na etapa classificatória.

Figure 11: The supergloss of artistic gymnastics[1]

---

[1] Translation: "Artistic gymnastics is a sport of formal precision in which gymnasts must present a routine composed of acrobatic and gymnastic elements in one of the apparatuses of the competition. In women's competition, gymnasts perform on four events: uneven bars, vault, floor and balance beam. In men's competition, gymnasts compete on vault and floor too, and also on the still rings, the horizontal bar, the parallel bars, and the pommel horse. A jury composed of 8 judges evaluate the gymnasts according to the level of difficulty of the routine (based on the value of the elements that make up each routine, established by the punctuation code) and also according to its execution (according to the quality and technical accuracy of the movements performed by the gymnasts). When performing the routine, gymnasts can make mistakes, which lead to score deduction, or perform combined movements or highly difficult movements, which lead to bonus points. These items determinate the gymnast routine score. The routines usually present an entry, a way of starting the presentation and contacting the apparatus, the execution of the elements of the routine, and dismount, the ending of the routine and the termination of contact with the apparatus. Different kinds of elements are performed by the gymnasts in acrobatics. Somersaults, pirouettes, dance jumps and supports are some of the elements which, performed in sequence, make up the routine. Present in Olympics since the first edition of the Modern Games, in Athens in 1886, artistic gymnastics competitions consist of individual all-around, team, or individual events. In the finals, only 8 gymnasts or teams that get the best scores in the qualifying round compete."
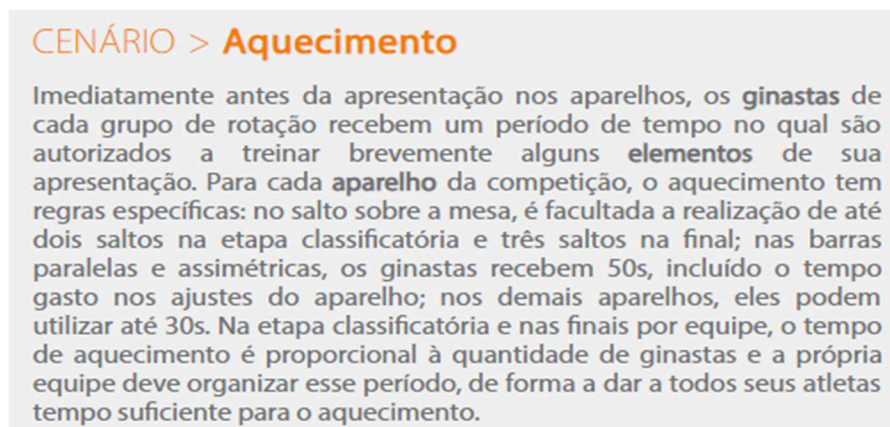
Figure 12: The gloss of one of artistic gymnastics frames[2]

The examples above show that the glosses of DO are strongly characterized by the use of encyclopaedic information and by their extended size, which are two important features that distinguish the DO glosses from the lexicographical definitions usually found in semasiologic, monolingual dictionaries. The first feature (the encyclopaedic information) brought to the DO compilation one of the biggest challenges faced by the SemanTec group during the writing of the glosses, impacting also on the second feature (the size of the glosses).

The next paragraphs approach this experience of writing the glosses. First and foremost, it is important to highlight that the distinction between linguistic knowledge and encyclopaedic knowledge has pervaded debates in Linguistics for a long time. A very important contribution from Cognitive Semantics to this discussion is the intensification of the idea that it is not always possible to make a rigid distinction between knowledge of language and knowledge of facts (see Evans & Green, 2006: 160-162; Riemer, 2010: 100-105; Geeraerts, 2010:222-224). As Riemer (2010: 104) postulates, "we know a variety of things about words and their denotation, and the greater the likelihood that a particular piece of this knowledge is shared between speaker and hearer, the greater the likelihood that it will determine the word's linguistic properties".

One of the consequences of this discussion to lexicography concerns the lexicographical definition, and, in particular, the content of definitions: how can lexicographers choose the best encyclopaedic information to define lexical items? If, on the one hand, "linguistic" information seems easier to be identified and chosen for the writing of definitions, on the other hand, encyclopaedic information corresponds to a larger

---

[2] Translation: "Immediately before the performance on the apparatus, the gymnasts in each rotation group are given a period of time in which they are allowed to briefly train some elements of their presentation. For each apparatus of the competition, the warm-ups have specific rules: for the vault, it is allowed to perform up to two vaults in the qualifying stage and three vaults in the final; for parallel and asymmetric bars, gymnasts have 50 seconds, including time spent adjusting the apparatus; for other apparatus, they can use up to 30 seconds. In the qualifying rounds and finals per team, the warm-up time is proportional to the number of gymnasts, and the teams must organize themselves in order to give all their athletes enough time to warm up."

amount of information, once it consists of our "knowledge of the world" (Matthews, 2007, s.v. encyclopaedic knowledge). This knowledge of the world represents an immeasurable amount of information; and it would be obviously impossible to allocate all the encyclopaedic information about a word in a single dictionary entry. Thus, when a lexicographer proposes to create encyclopaedic definitions for dictionary entries – whether brief definitions of printed semasiological dictionaries or longer definitions, such as the definitions of DO – this lexicographer will always face the challenge of choosing the most appropriate information to describe lexical units.

Let us take the example of football. Which information would be indispensable to describe its meaning? The fact that it is a sport in which players use their feet? That the objective is scoring goals? That the teams have supporters? That the games take place at stadiums? That the match is played by two teams? That the teams are composed by eleven players? That each match is divided into a first and a second half of 45 minutes each? That between the first and the second half there is a break of 30 minutes? That there is an official football World Cup? That Pelé is considered the king of football? We emphasize that we are not even trying to separate linguistic from encyclopaedic knowledge – we are just trying to list what is essential in the definition of the word *football*.

In the context of the DO, without having a methodology that could guide the lexicographers to choose the most adequate information for the description of the sports, each editor found their own way to describe the sports they were responsible for. They used especially their linguistics intuition based on the studies previously developed about the sports. At the end, the editors compiled a group of glosses which could meet the demands of the DO users, although there were significant differences between them, especially because of the size and kind of information presented. The two examples below demonstrate this:
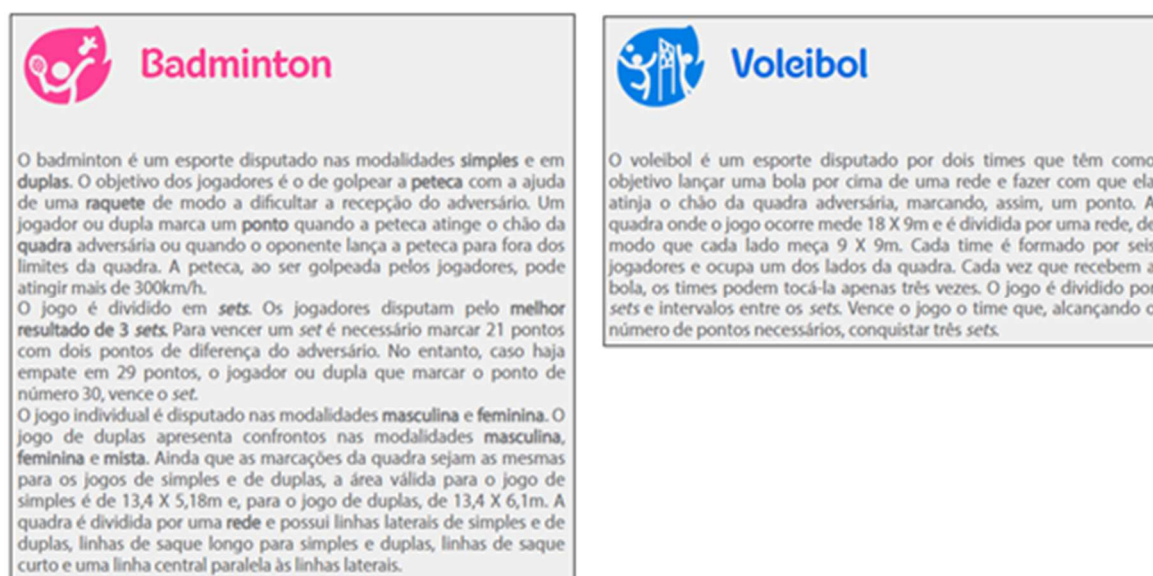

Figure 13: Size difference between two glosses

The images above present the size difference between two superglosses. Once they define different sports, it would be expected the two could diverge from each other with regard to length, especially because some sports may require specific explanations about specific features, while other sports may not. Even so, we believe that it would be possible to create a methodology for the writing of the superglosses, presenting them in a more standardized form, especially in terms of size and content.

We have put this into practice during the compilation of the Dicionário Paraolímpico, the most recent dictionary produced by the SemanTec group that is currently under construction. Dicionário Paraolímpico will present the same lexicographic structure as the DO and will also have Frame Semantics as a guideline. This dictionary has benefited from all the expertise acquired by the group during the compilation of the DO, which has been helping the group to reflect on new strategies to solve some challenges such as the writing of the glosses.

The methodology for the writing of the glosses of Dicionário Paraolímpico is currently under development. It proposes to split the gloss into two parts. The first part is intended to have the form of an intensional definition, which enumerates a set of important features of the Paralympic sports. To construct this part, we base our work on a study that proposes a classification of sports (Gonzalez, 2004). In this study, Gonzalez (2004) classifies the sports based on four parameters that he calls "relation to the opponent", "relation to the objective", "relation to the partner" and "relation to the environment". Considering this division, the first part of the gloss will present the information below (we added one more parameter, the "objective"):

1) Kind of sport (relation to the objective): translation / fight / field and bat or court / split court or wall / by demarcation / aesthetic or technical combinatory / precision or target

2) Relation to the partner: individual / collective

3) Objective

4) Relation to the opponent: interaction with the opponent or direct opposition to the opponent / no interaction and no direct opposition

5) Relation to the environment: stability / no stability

The second part of the gloss describes some specifics of the Paralympic competition, opposing, if possible, the Paralympic sport to its Olympic counterpart. In this part of the gloss, we intend to include an extensional format of definition that provides the user with encyclopaedic information about the Paralympic sports. Putting this methodology into practice, we have developed the following template to guide the writing of Paralympic sports' glosses:

PART 1: _____(name of the sport)_____ é um esporte de/do tipo _____(1)_____ disputado/que pode ser disputado _____(2)_____ cujo objetivo é _____(3)_____ [descrição da sequência do ato esportivo]. No(a) _____(name of the sport)_____, a relação com o adversário ocorre de maneira _____(4)_____ através de [descrever relação entre atletas no ato da competição]. O _____ (name of the sport)_____ é praticado em [descrever ambiente], ambiente que oferece/não oferece _____5_____ para o atleta.

PART 2: specificities of Paralympic competition and the differences between the Paralympic sport and its Olympic counterpart.

The following gloss is an example of application of this template to a Paralympic sport – football 5-a-side:

O futebol de 5 é um esporte **de quadra** disputado **coletivamente** cujo objetivo **de cada equipe é marcar gols na área adversária**. No **futebol de 5**, a relação com o adversário ocorre por **oposição direta** através de **disputas de bola, dribles, passes e chutes a gol**. O **futebol de 5** é **geralmente** praticado **em quadras adaptadas de futebol de salão, podendo também acontecer em campos de grama sintética**, ambientes que oferecem **estabilidade** para o atleta. **Em relação a sua contraparte olímpica, o futebol de cinco diferencia-se por ser disputado por atletas cegos, que utilizam vendas nos olhos para garantir condições iguais a todos os participantes. A bola da partida possui guizos internos para que os jogadores possam localizá-la e a quadra possui bandas junto às linhas laterais, para evitar que a bola saia. Durante a partida, existe um guia, que recebe o nome de chamador, que fica atrás do gol para orientar os jogadores em relação ao seu posicionamento em campo e chutes a gol. As partidas de futebol de 5 acontecem de maneira silenciosa; a torcida só tem permissão de se manifestar quando acontecem gols.**[3]

---

[3] Translation: "5-a-side football is an indoor sport played collectively whose purpose is to score goals in the opposing area. In a 5-a-side football match, the relationship of opponents occurs by direct opposition through ball disputes, dribbling, passes and goal shots. 5-a-side football is usually played on courts adapted from indoor soccer and may also take place on synthetic grass, places that offer stability for the athletes. Differently from its Olympic counterpart, 5-a-side football is played by blind athletes who use blindfolds to ensure equal conditions to all participants. The ball has bells inside to aid the players in their movements, and the court has bands along the lines to prevent the ball from coming out. During the match there is a guide, who receives the name of caller and stands behind the goal to guide the players' positions and shots. 5-a-side football has quiet matches; the public is allowed to cheer only when a goal is scored."

5. Final considerations

This paper presented an overview of the challenges and difficulties faced in the development of the Dicionário Olímpico. In the previous pages, we presented some problems we faced during our work and how we dealt with some of these issues. As many studies have shown, Cognitive Linguistics and Frame Semantics have proved to be important theoretical frameworks for lexicography (especially for online dictionaries). Considering this potential, one of the biggest challenges of Cognitive Lexicography is to build its own methods to convert the principles of a cognitive theory of language into tools for dictionary making. The lexicographic products presented in this paper integrate this enterprise.

6. Acknowledgements

7. References

Atkins, S. (2002) Then and now: competence and performance in 35 years of lexicography. In A. Braasch, & C. Povlsen (eds.) *Proceedings of the Tenth EURALEX International Congress*. Copenhagen: Center for Sprogteknologi, pp. 247-272.

Atkins, S., Fillmore, C. J. & Johnson, C. R. (2003). Fuzzy SF: Lexicographic Relevance: Selecting Information from Corpus Evidence. *International Journal of Lexicography*, 16(3), pp. 251–281.

Atkins, S., Rundell, M. & Sato, H. (2003). The contribution of FrameNet to Practical Lexicography. *International Journal of Lexicography*, 16(3), pp. 333–357.

Chishman, R. et al. (2015). The relevance of the Sketch Engine software to build Field - Football Expressions Dictionary. *Revista de Estudos da Linguagem*, 23, pp. 769-796.

Chishman, R. et al. (2017) Dicionário Olímpico: a semântica de frames encontra a lexicografia eletrônica. In M. J. B. Finatto et al (eds.) *Linguística de Corpus: Perspectivas*. Porto Alegre: Instituto de Letras - UFRGS, pp. 265-298.

Evans, V. & Green, M. (2006). *Cognitive Linguistics*: an introduction. Edinburgh: Edinburgh University Press.

Fillmore, C. J. (1982). Frame Semantics. In The Linguistics Society of Korea (ed.) *Linguistics in the Morning Calm*. Seoul: Hansinh Publishing Co., pp. 111–137.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2), pp. 222-254.

Fillmore, C. J. & Atkins, S. (1992). Toward a Frame-based Lexicon: The Semantics of RISK and its Neighbors. In A. Lehrer & E. Kittay (eds.) *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*. Hillsdale: Erlbaum,

pp. 75-102.

Fillmore, C. J. & Baker, C. (2009). A frames approach to semantic analysis. In B. Heine & H. Narrog (eds.) *The Oxford Handbook of Linguistic Analysis.* New York: Oxford University Press, pp. 313-339.

Geeraerts, D. (2010). *Theories of Lexical Semantics.* New York: Oxford University Press.

Gonzalez, F. J. (2004). Sistema de classificação de esportes com base nos critérios: cooperação, interação com o adversário, ambiente, desempenho comparado e objetivos táticos da ação. *Lecturas: Educación Física y Deportes*, 10(71).

Hartmann, R. R. K. & James, G. (2002). *Dictionary of lexicography.* 2.ed. London/ New York: Routledge.

Matthews, P. H. (2007). *Oxford Concise Dictionary of Linguistics.* 2nd ed. Oxford: Oxford University Press.

Ostermann, C. (2012). Cognitive lexicography of emotion terms. In R. V. Fjeld. & J. M. Torjusen (eds) *Proceedings of the 15th EURALEX International Congress.* Oslo: Department of Linguistics and Scandinavian Studies/University of Oslo, pp. 493-501.

Ostermann, C. (2016). *Cognitive Lexicography: a new approach to lexicography making use of Cognitive Semantics.* Berlin; New York: Mouton de Gruyter.

Riemer, N. (2010). *Introducing Semantics.* New York: Cambridge University Press.

Santos, A. & Chishman, R. (2015). O papel da Semântica de Frames na construção de um recurso dicionarístico: a organização lexicográfica do FIELD - Dicionário de Expressões do Futebol. *Revista da ABRALIN*, 14, pp. 433-468.