

A good TACTIC for lexicographical work: football terms encoded in TEI Lex-0

Ana Salgado, Rute Costa

NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa
Avenida de Berna, 26-C 1069-061 Lisboa, Portugal
anasalgado@campus.fcsh.unl.pt
rute.costa@fcsh.unl.pt
<https://clunl.fcsh.unl.pt/equipa/ana-salgado/>
<https://clunl.fcsh.unl.pt/equipa/rute-costa/>

Abstract. Terms are a significant part of lexicographical nomenclatures in general language dictionaries. In this paper, we focus on how football terms are treated in three Academy Dictionaries – Portuguese, French, and Spanish – and draw some conclusions about the lexicographical decisions taken in the three languages. After identifying every position football players can have on the field, we verify whether the dictionaries above include these terms. We propose the TEI encoding of the term “defesa” (defence), which designates a position occupied by football players on the field. Bearing in mind concepts such as reusability and interoperability, we intend to present: 1) a comparison of football terms in the three dictionaries; 2) TEI Lex-0 dictionary encoding, a streamlined standard to facilitate interoperability; 3) a consistent TEI modelling and description of the microstructural elements of lexicographical entries. In the end, we draw some conclusions.

1. Introduction

Lexicography is conceived as a field that deals mainly with lexical units (words) but also with specialised lexical units (terms). While Lexicography and Terminology are two different scientific disciplines – with different theoretical-epistemological backgrounds – they have in common the fact that they deal with terms, albeit with different aims more often than not. This means

that working in Terminology and Lexicography requires different approaches since the social, cultural or economic purposes are not the same.

Lexicographers follow mostly a semasiological perspective (from words to senses), and terminologists, mostly concept-oriented, combine conceptual organisation and linguistic analysis where the definition of the concept is central with the view to reduce linguistic ambiguities. In 2013, Costa stated that Terminology and Lexicography should be seen as complementary regarding the methods they use. Bowker (2018, p. 149), arguing for the relation between these fields, sees advantages in the fact that “lexicographers and terminologists continue to work together to tackle new challenges and embrace new opportunities”.

Getting to know the domain and subsequently organising it are two requisite activities for a rapid and systematic identification of the basic concepts, which will result in a better description of the terminology. Bearing in mind that we are working with a specialised domain, the intervention of the expert is necessary to aid in the task of organising knowledge and to validate the descriptions and definitions of terms (Silva & Costa, 2019). This facilitates a more accurate encoding by allowing a tidier classification of the data depending on each element.

The primary purpose of this paper is to show how football terms are treated in the three Academy dictionaries that comprise our lexicographical corpus (i.e. Portuguese, French, and Spanish languages)¹ (section 2). After identifying every position football players can have on the field, we confirmed whether the Academy dictionaries include these football terms (section 3). A contrastive study of the corpus allowed us to observe the data and draw conclusions about the lexicographical work performed in the three languages, gaining some insights into the lexicographical decisions that have been made in the three different dictionaries.² Subsequently, we applied and tested the latest version of the TEI Lex-0 guidelines by representing the term “defesa” (defence), which designates a position occupied by players on the field (section 4). Aiming at providing guidelines for a structurally organised and consist-

1 DLPC = *Dicionário da Língua Portuguesa Contemporânea*, 2001, Academia das Ciências de Lisboa; DAF = *Dictionnaire de l'Académie Française*, 2019, Académie Française, <http://www.dictionnaire-academie.fr/>; DLE = *Diccionario de la Lengua Española*, 2019, Real Academia Española, www.rae.es/rae.

2 This research is part of the PhD project of the first author that intends to design and test a set of methodological guidelines that can facilitate the inclusion of terms in a general language dictionary by systematising information regarding a single domain.

ent processing of lexicographical information, and bearing in mind concepts such as reusability and interoperability, we intend to present: 1) arguments in favour of using TEI Lex-0 dictionary encoding, a streamlined standard to facilitate interoperability; and 2) a consistent encoding and description of the microstructural elements of lexicographical entries, exemplifying the representation of lexicographical content using the term “defesa” (defence).

2. The lexicographical corpus

Recognising the importance of national academies that aim to create a dictionary in order to preserve the language, we decided to build a lexicographical corpus consisting of Academy works. The Academies of Sciences have undertaken dictionaries that are considered official authorities on the usages and the vocabulary of a language.

Our lexicographical corpus is comprised of three dictionaries published by different Academies: the *Dicionário da Língua Portuguesa Contemporânea* by the Academia das Ciências de Lisboa (DLPC), the *Diccionario de la Lengua Española* by the Real Academia Española (DLE), and the *Dictionnaire de l'Académie Française* by the Académie Française (DAF). These are general language contemporary dictionaries with printed editions and a descriptive nature with a normative concern addressed to a vast audience. The reason why we decided to create a contrastive corpus is justified by the fact that although the languages are different and the dictionaries themselves are also different, they share similar problems.

The content of these dictionaries is written in the languages of origin of each institution, i.e. Portuguese, Spanish, and French. These three dictionaries are available online³, accessible for free, and are updated continuously.

In Portugal, despite the successive attempts of the Academy, only in 2001, under the coordination of Malaca Casteleiro, did the Academia das Ciências de Lisboa publish a complete dictionary (from A to Z) for the first time in a two-volume paper version: *Dicionário da Língua Portuguesa Contemporânea* (DLPC). A new digital version – a task that has been undertaken by a team working in Natural Language Processing (NLP) at the University of Minho

3 The DLPC is currently only available in-house. However, as the coordinator of the new dictionary of the Academia das Ciências de Lisboa, the first author is making steps towards its public availability.

– is the basis of the ongoing review of this dictionary, which now counts with the participation of NOVA CLUNL.⁴

The *Diccionario de la Lengua Española* is the widest normative dictionary of Spanish, published and created by the RAE. The most recent edition is its 24th, and it has been online since 2005.

The official dictionary of the French language, known as the *Dictionnaire de l'Académie Française* (DAF), served as a model for other dictionaries for many years, and the new version was made available on 7 February 2019, integrating the 9th edition in progress, from letters A to S.

3. Football domain

Football is the domain we have chosen to test the proposal for a set of future methodological guidelines for the lexicographical processing of terms. Our interest in football arises from the fact that it has been the most popular sport on the planet since the end of the 19th century; a sport with worldwide expansion via different societies in every continent. It is estimated that 250 million people are directly involved in football and that 1.4 billion people in the world have some interest in football (Morris, 1985).

We also aim to observe popularisation of the terms, namely the transition of a term into the vocabulary of everyday language. Many football terms have been adopted in everyday language, such as “canto” (corner) or “defesa” (defence) for instance.

This sport is also often referred to as 11-player football because it is played between two teams of 11 players each as seen on each of the definitions of the term in the three dictionaries that comprise our corpus (Fig. 1): “onze jogadores” (DLPC), “onze joueurs” (DAF), “once jugadores” (DLE):

4 The Natural Language Processing group of the Computer Science Department of the University of Minho has been developing the technological support of the new digital edition of the DACL, counting on the participation of Alberto Simões from IPCA (Instituto Politécnico do Cávado e do Ave), responsible for the technological support, José João Almeida, and the consultancy of Álvaro Iriarte Sanromán, both from the University of Minho. The participation of NOVA CLUNL (Linguistic Research Center of NOVA University of Lisbon) is related to its transition into the TEI LEX-0 format.

futebol [futib3l]. *s. m.* (Do ingl. *football*). *Desp.* 1. Modalidade desportiva em que jogam duas equipas, de onze jogadores cada uma, que procuram controlar e introduzir a bola na baliza da equipa adversária. *O futebol é a modalidade desportiva mais popularizada em Portugal. Jogo, desajo de +; clube, equipa, jogador de +; campo de +; jogar +.*

fútbol

Tb. futbol.

Del ingl. football.

1. *m.* Juego entre dos equipos de once jugadores cada uno, cuyo objetivo es hacer entrar en la portería contraria un balón que no puede ser tocado con las manos ni con los brazos, salvo por el portero en su área de meta.

FOOTBALL (se prononce *foutbo*) **nom masculin**

xviii^e siècle. Mot anglais composé de foot, « pied », et ball, « balle, ballon ».

Désignait initialement tout jeu de ballon opposant deux équipes. Aujourd'hui, est réservé au jeu mettant aux prises deux équipes de onze joueurs qui doivent faire entrer un ballon, sans le toucher de la main ou du bras, dans les buts défendus par le camp adverse. *Un terrain de football. Un joueur, un club de football. Football amateur, professionnel. La Coupe de France, la Coupe du monde de football.*

FIG. 1 – Entry “*futebol/football/fútbol*” (DLPC, DAF, DLE)

3.1. The positions of football players on the field

The 11 football players occupy specific positions on the field which are connected with specific terms (Fig. 2):



FIG. 2 – *Football players occupy different positions on the field*

Abbrev.	Portuguese	French	Spanish	English <small>(literal translation)</small>	DLPC	DAF	DLE
GUARDA-REDES (gardien de but, portero, goalkeeper)							
GR	guarda-redes	gardien de but	portero	goalkeeper	✓	✓	✓
DEFESA (defense, defensa, defender)							
LD	lateral direito	arrière latéral droit	lateral derecho	right-back	—	—	—
LE	lateral esquerdo	arrière latéral gauche	lateral izquierdo	left-back	—	—	—
DC	defesa central	défenseur central	defensa central	centre-back	—	—	—
LB	libero	Libéro	libero	sweeper	✓	—	✓
MEIO-CAMPO (centro del campo, milieu du terrain, midfield)							
MD*	médio defensivo	milieu défensif	volante de corte	defensive midfielder	—	—	—
MD	médio direito	milieu droit	volante externo derecho	right midfielder	—	—	—
ME	médio esquerdo	milieu gauche	volante externo izquierdo	wide midfielder	—	—	—
MC	médio-centro	milieu central	volante central	centre midfielder	—	—	—
MO	médio ofensivo	milieu offensif	volante ofensivo	attacking midfielder	—	—	—
ATAQUE (ataque, attaque, attack)							
AV	avançado	avant	delantero	centre forward	✓	—	✓
SA	segundo-avançado	deuxième avant	segundo delantero	second striker	—	—	—
PL	ponta de lança	buteur	punta	striker	✓	—	✓
ED	extremo-direito	aillier droit	extremo derecho	right winger	—	—	—
EE	extremo-esquerdo	aillier gauche	extremo izquierdo	left winger	—	—	—

Tab. 1 – *Terms referring to positions occupied by football players on the field*

Looking at Tab. 1, we can see that only the term “goalkeeper” is recorded in all the dictionaries. Most terms that designate the positions of the players are not recorded in our dictionaries – e.g., “right-back”, “left-back”, “centre-back”, “right winger”, and “left winger”. We may argue that this happens because we are dealing with polylexical units, such as “left back”, and not just with monolexical units, such as “back” in English, “lateral” in Portuguese, “latéral” in French, and “lateral” in Spanish. In consequence, we decided to search for these units in our lexicographical corpus. The unit “lateral”, when related to football, figures in the DLPC (“Fut. Jogador que actua junto da linha lateral do campo.”⁶) and in the DLE (“Dicho de un futbolista o de un jugador

6 Player acting near the sideline.

de otros deportes: Que actúa junto a las bandas del terreno de juego con funciones generalmente defensivas.”⁷), but is absent from the DAF.

To avoid such inconstancies, a terminological approach to the domain would be of major help. Building a concept system by identifying the relations between the concepts that embody the positions occupied by football players would allow the lexicographers to compile all the terms designating them. A conceptual approach to domains prevents lexicographers from missing essential terms of a terminology.

A term that is included in all these dictionaries, “goalkeeper”, raises some controversial questions. Although the DLPC uses “Fut.” (football) as a domain label listed in the abbreviation list, in the case of “guarda-redes”, the domain label used is “Desp.” (sports) (“Desp. Jogador que, no jogo do futebol, andebol, hóquei... ocupa o último posto de defesa, entre os postes da baliza, tentando impedir a marcação de golos”⁸). This happens because the ‘definition’ presented above is not only related to the football domain, but includes other sports. In the DLE, “portero” is not identified by any label (“Jugador que en algunos deportes defiende la portería de su bando”⁹). Finally, the DAF use the “Sports” label (“SPORTS. Gardien de but, joueur assurant la défense des buts dans certains jeux de ballon”¹⁰).

It seems clear that the DLPC and the DAF distance themselves from the DLE by using the domain label to differentiate meanings or contextualize them, merely specifying the domain of the meaning. In fact, any criterium can be validated as long as it is applied uniformly.

4. Text Encoding Initiative (TEI)

In order to make content interoperable and reusable, we decided to follow the Guidelines of the Text Encoding Initiative (TEI), which is a *de facto* standard in digital edition. These guidelines provide a dictionary module (TEI Consortium¹¹) and have been used in numerous dictionary projects for

7 Said of a football player or a player of other sports: One that acts alongside the sidelines with generally defensive functions.
8 Player who, in football, handball, hockey..., occupies the last defense position between the goal posts, trying to prevent the scoring of goals.
9 Player who in some sports defends the goal of their side.
10 Goalkeeper, player defending goals in certain ball games.
11 <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>

born-digital lexicographical data (Budin *et al.*, 2012) or retrodigitised projects (Bohbot *et al.*, 2018).

However, since the TEI Guidelines present numerous and flexible encoding possibilities, a new version specifically for dictionary encoding is now being discussed –TEI Lex-0¹² (Romary and Tasovac, 2018; Bański *et al.*, 2017). We have chosen to test this target format, a streamlined standard to facilitate interoperability and some best-practice guidelines for NLP purposes.

TEI Lex-0 will not replace the “Dictionaries” chapter of the TEI Guidelines; instead it is being discussed as a target format that will standardise the existing heterogeneously encoded lexical resources and is being tested by numerous language dictionaries (Salgado *et al.*, 2019b).

4.1. TEI Lex-0 encoding

The application of TEI Lex-0 will be demonstrated with samples of the term “defence” (Figures 4, 5, and 6) of the DLPC, the DAF, and the DLE. A consistent encoding and description of the microstructural components of the lexicographical article was the first step of this codification. Looking at the three dictionaries, the entries “defesa” (Fig. 4), “défense” (Fig. 5), and “defensa” (Fig. 6) include the same following elements: headword; etymology; part-of-speech (in the DLE this information is given in sense); a series of numbered meanings (polysemy). Depending on the criteria adopted in each of the dictionaries, we can also have domain labelling and other type of usage information, synonyms, collocations, and examples of usage.

12 TEI Lex-0 — A baseline encoding for lexicographical data: <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>

defesa¹ [difẽns]. s. f. ¹(Do lat. *defensa*). 1. Acção de proteger ou proteger-se; acto ou efeito de defender ou de defender-se. ≠ *ataque*. *A defesa de um país. Instinto de si; reflexo de «.* **legítima**² *defesa.* 2. O que serve para proteger, defender. *O guarda-chuva era a sua única defesa contra o aguaceiro. A trincheira era a única defesa dos soldados. Era forte já foi a principal defesa da ilha. Arma de «.* *linha de si; meio de «.* **defesa nacional**, conjunto de meios destinados a assegurar a integridade territorial de um país contra possíveis ataques estrangeiros. **mecanismo**³ *de defesa.* 3. Acto de patrocinar, de sustentar, de defender uma ideia, uma causa. *A defesa da liberdade. A defesa de uma tese. A defesa de uma ideologia. A defesa de um ideal.* **defesa de tese**, acto solene em que um candidato a determinado grau académico apresenta um trabalho inédito e argumenta em sua defesa perante um júri. 4. *Dir.* Advogado que defende um réu em juízo. *O juiz deu a palavra à defesa. A defesa apresentou as alegações finais. Advogado de «.* 5. *Dir.* Acto de justificar, defender ou defender-se um réu em tribunal contra uma acusação, expondo factos e produzindo provas em seu favor. = *acusação. A defesa e a acusação.* 6. *Dir.* Conjunto de procedimentos adoptados pelo réu ou pelo seu advogado de defesa para refutar as acusações. *O advogado preparou cuidadosamente a defesa do arguido.* 7. Acto de impedir, de proibir, = *impedimento, interdição, proibição.* 8. *Us. pl. Zool.* Chifres de animais. = *chifres, cornos, nariz.* 9. *Us. pl. Zool.* Dente muito saliente em alguns animais. *As defesas do elefante. As defesas do javali. As morsas deprimam-se, utilizando as defesas.* 10. Preservativo. 11. *Dep.* Intercepção da bola pelo guarda-redes, impedindo-a de entrar na baliza. *O guarda-redes fez uma defesa incrível.* 12. *Dep.* Acção de contrariar o ataque, a investida do adversário; acto ou efeito de defender. = *ataque. «Errado é jogar à defesa.» (DN, 26.10.1988).* 13. *Dep.* Conjunto de jogadores que têm como função contrariar o ataque do adversário, actuando na parte recuada do meio campo da sua equipa. *A equipa adversária deparou com uma defesa intransponível.* 14. *Equit.* Resistência oposta por um cavalo à transposição de um obstáculo ou para se subtrair à acção do cavaleiro. 15. *Region. (Alenx.)* Grande herdade ou propriedade rústica. **em defesa de**, *loc. prep.*, em apoio de; em favor de. *O governo tomou medidas em defesa do meio ambiente. O sindicato actuou em defesa dos direitos dos trabalhadores.* **em legítima defesa**, *loc. adv.*, em reacção a uma agressão ilícita, actual ou iminente. *Disparou em legítima defesa. Meteu em legítima defesa. jogar à defesa.* 1. *Dir.* Procurar defender a sua baliza, sem atacar, sem procurar marcar golos. 2. Não se expor. *Não consegue ser espontânea por estar permanentemente a jogar à defesa; não ter defesa (possível).* *Fam.*, não ter qualquer justificação para a prática de um acto considerado errado e reprovável. **defesa**² [difẽns]. s. m. (Do lat. *defensus*). *Dep.* Jogador de futebol ou de outros desportos, que actua na parte recuada do meio campo da sua equipa. *«ficou na frente, não apenas para dar continuidade aos lances mas também para 'brincar' ou defesa contrária.» (DN, 27.10.1988).*

FIG. 4 – Entry “defesa” (DLPC)

I. DÉFENSE nom féminin

XIII^e siècle. XIX^e siècle, comme terme de droit, au sens de « la partie qui se défend ». Emprunté du bas latin *defensa*, « défense », participe passé féminin substantivé de *defendere*, « défendre ».

1. Action de défendre quelqu'un ou de se défendre contre une attaque. *Venir, courir à la défense* [...]
3. Par analogie. MÉDECINE. *Défense musculaire*, contraction douloureuse des muscles (s'emploie surtout à propos de la paroi abdominale, où la défense musculaire témoigne le plus souvent d'une inflammation péritonéale). – PHYSIOLOGIE. *Les défenses de l'organisme*, ses moyens de résistance à l'invasion microbienne, virale, parasitaire, etc. *L'immunologie étudie le système de défense de l'organisme*. – PSYCHOLOGIE. *Une réaction de défense. L'instinct de défense. Mécanismes de défense*, moyens par lesquels le sujet cherche inconsciemment à se protéger. – JEUX DE BALLON. Action de défendre les buts d'une équipe, de résister aux attaques de l'équipe adverse et, par extension, en parlant d'un sportif, action de s'opposer aux offensives de l'adversaire. Par métonymie. *La défense, les lignes arrière d'une équipe, chargées de défendre les buts*. – ÉQUITATION. *Les défenses d'un cheval*, les mouvements par lesquels un cheval se dérobe à l'action de son cavalier. – MARINE. *Défenses*, morceaux de bois, tampons de cordage, pneus usagés, cylindres de matière plastique disposés le long du bord d'un navire afin d'amortir les chocs ou d'empêcher les frottements (on dit aussi *Pare-battage*).

FIG. 5 – Entry “*défense*” (DAF)

defensa

Del lat. tardío *defensa*.

1. f. Acción y efecto de defender o defenderse.
2. f. Arma, instrumento u otra cosa con que alguien se defiende en un peligro.
3. f. Amparo, protección, socorro.
4. f. Obra de fortificación que sirve para defender una plaza, un campamento, etc. U. m. en pl.
5. f. Jugada del tresillo en la que un jugador sustituye en sus derechos y deberes al hombre que rinde la jugada.
6. f. Mecanismo natural por el que un organismo se protege de agresiones externas. U. t. en pl. con el mismo significado que en sing.
7. f. **línea defensiva.**
8. f. En ajedrez, conjunto de jugadas previstas que constituyen una determinada estrategia.
9. f. Razón o motivo que se alega en juicio para contradecir o desvirtuar la pretensión del demandante.
10. f. Abogado defensor del litigante o del reo. U. m. en pl. cuando hay varios reos en el mismo juicio.
11. f. *Cuba, Méx., R. Dom. y Ur. parachoques.*
12. f. pl. Colmillos del elefante, cuernos del toro, etc.
13. f. pl. *Mar.* Conjunto de objetos que se cuelgan del costado de la embarcación para que esta no roce o golpee contra el muelle u otra embarcación.
14. m. y f. Cada uno de los jugadores que forman la línea defensiva.

FIG. 6 – Entry “*defensa*” (DLE)

Focusing on the form element that contains the orthographic, phonetic, and grammatical information of the entry (headword), all the elements of the entries are distinguished as clearly as possible:

<pre> <entry xml:lang="pt" xml:id="defesa_1" n="1"> <form type="lemma"> <orth>defesa</orth> <pron>difɛzɐ</pron> </form> <gramGrp> <gram type="pos" norm="NOUN">s.</ gram> <gram type="gen">f.</gram> </gramGrp> <etym> [...]</pre>	<pre> <entry xml:lang="fr" xml:id="d�fense"> <form type="lemma"> <orth>d�fense</orth> </form> <gramGrp> <gram type="- pos" norm="- NOUN">nom</gram> <gram type="gen">f�minin</ gram> </gramGrp> <etym> [...]</pre>	<pre> <entry xml:lang="es" xml:id="defensa"> <form type="lemma"> <orth>defensa</orth> </form> <etym> [...]</pre>
--	--	--

FIG. 7 – Entry “defesa/d fense/defensa” (DLPC, DAF, DLE) encoded in TEI Lex-0

From these encoding examples, we highlight that the TEI Lex-0 schema only uses entry, the basic element of the dictionary’s microstructure, once we have constrained the general structure of a lexical entry – in our schema, entryFree, superEntry and re (related entry) of the current Guidelines are not used. In FIG. 7, entry contains the following elements – form, etym, and gramGrp – and requires the attributes @xml:id and @xml:lang, the appropriate language code (BCP 47) – “pt” (DLPC,; “fr” (DAF), or “es” (DLE). Dictionary entries always start with a lemma (headword), a canonical form. The lemma form is encoded as form @type=“lemma”.

To specify the morpho-syntactic properties of the entry, we use gramGrp, which includes the part-of-speech of the entry (pos) and further specifications, such as the morphological gender (gen) of a lexical item. In

the example, for part-of-speech we use the universal dependencies tagset.¹³ In addition, we highlight the fact that the DLPC has also phonetic transcription (pron) and a numerical index (number), and, in the DLE, the part of speech (a noun) is inferred from the attribution of the female gender and is marked in sense.

13 In the ELEXIS context, a @norm attribute is mandatory to specify a normalised (UD) part of speech value. For more details: <https://universaldependencies.org/u/pos/>

<pre> <sense xml:id="de- fesa_11" n="11"> <usg type="domain">- Desp.</usg> <def>Intercepção da bola pelo guarda-redes, impedindo-a de entrar na baliza.</def> <cit type="example"> <quote>O guarda-redes fez uma defesa incrível.</quote> </cit> </sense> <sense xml:id="def- esa_12" n="12"> <usg type="domain"> Desp.</usg> <def>Acção de contra- riar o ataque, a investida do adversário; acto ou efeito de defender.</ def> <xr type="antonymy"> <ref type="sense">ata- que</ref> </xr> <cit type="example"> <quote>«Errado é jogar à defesa.»</quote> <bibl> <title>DN</title> <date>26.10.1998</ date></bibl> </cit> </sense> [...]</pre>	<pre> <sense xml:id=>dé- fense_3.4» n=>3»> <usg type=>domain>>- JEUX DE BALLON.</ usg> <def>Action de défendre les buts d'une équipe, de résister aux attaques de l'équipe adverse et, par exten- sion, en parlant d'un sportif, action de s'op- poser aux offensives de l'adversaire.</def> </sense> <sense xml:id="de- fensa_3.5" n="3"> <usg type=>mea- ningType>>Par métony- mie.</usg> <cit type=>example>> <quote>La défense</ quote> <title></title> <def>les lignes arrière d'une équipe, chargées de défendre les buts.</ def> </sense> [...]</pre>	<pre> <sense xml:id="de- fensa_7" n="7"> <gramGrp> <gram type="gen">f.</ gram> </gramGrp> línea defensiva. </sense> [...]</pre> <pre> <sense xml:id="de- fensa_14" n="14"> <gramGrp> <gram type="gen">m.</ gram> <lbl>y</lbl> <gram type="gen">f.</ gram> </gramGrp> <def>Cada uno de los jugadores que forman la línea defensiva.</def> </sense> [...]</pre>
--	--	---

FIG. 8 – Entry “defesa/défense/defensa” (DLPC, DAF, DLE) encoded in TEI Lex-0

When dealing with football terms, the associated lexical items belong to a specific domain. This usage information is usually identified in dictionaries through domain labelling (Salgado and Costa, 2019; Salgado *et al.*, 2019a). In FIG. 4, senses 11, 12, and 13 are marked with “Desp.” (the abbreviated form of “Desporto” [Sports]). In FIG. 5, the sense related to football is marked with the expanded form “JEUX DE BALLON”. Lastly, in FIG. 6, senses number 7 and 14, which refer to football, are not annotated. In TEI-style encoding, domain is a marker that identifies the specialised field of a lexical item – in TEI Lex-0 we use the element *usg* to mark usage information such as this (Fig. 8). Looking at other examples from TAB. 1, you can detect inconsistencies, since, for example, in the DLPC, both sports and football domains are used (e.g. “defesa-direito” is marked with the football domain, while “ponta-direita”, a synonym of “extremo-direito”, again a position occupied by a football player on the field, is marked as sports). Finally, we will also highlight multiword expressions, such as “jogar à defesa” (to play defence), which are treated as entries. We argue that this information should be standardised to a single value.

5. Concluding remarks

We want to conclude with four final remarks:

1. Our research has strictly lexicographical purposes, using Terminology’s methodologies as a contribution to the definition of guidelines and a methodology for the selection, inclusion, and processing of terms in general language dictionaries, namely the Academy Dictionaries we have analysed here, proposing a new dictionary model that, in a harmonised and balanced way, combines lexicographical methodologies and terminological methodologies.
2. Combining conceptual organisation and linguistic analysis is a necessity. Getting to know the domain and subsequently organising it are two requisite activities for a rapid and systematic identification of the basic concepts, which will result in a better description of the lexicon. This facilitates encoding by allowing a tidier classification of the data depending on each element, such as entry or sense, or *@type* “domain”> attribute. Bearing in mind that we are working with specialised domains, the intervention of the expert is necessary to aid in the task of organising knowledge, which will result in more accurate encoding.
3. By examining the codification of the term “defesa”, we have confirmed that the Lex-0 TEI meets our research needs. After encoding the

microstructure of the dictionary, we can guarantee the interoperability and reusability of the information. The advantage of applying TEI Lex-0 lies in the fact that lexicographers and terminologists alike are currently making efforts to apply TEI to the ongoing revision of ISO LMF (Romary, 2015). Given its (still) non-standard nature, it can be changed to accommodate relevant dictionary structures. We intend to demonstrate that the results obtained are useful for computational lexical encoding and may serve the purpose of NLP.

4. Finding that the dictionaries that make up our lexicographical corpus share common problems regarding the example we worked with (“defesa”) has led us to suggest that it would be interesting to present identical solutions for all of them. The reason why we decided to create a contrastive corpus is justified by the fact that although the languages are different and the dictionaries themselves are also different, they share similar problems. A contrastive analysis of the three dictionaries allowed us to confirm that. The solutions we have presented for the Portuguese example (DLPC) would be replicable in dictionaries of other languages since TEI is a recommendation for standardisation. An agreement between academies and other institutions would be desirable to systematise and optimise a new type of Lexicography that can provide a better representation of the entire European lexicographical heritage.

Acknowledgements

Research financed by Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia as part of the project Centro de Linguística da Universidade NOVA de Lisboa – UID/LIN/03213/2020, and by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731015 (ELEXIS).

References

- Bański, P., Bowers, J., and Erjavec, T. 2017. “TEI Lex-0 guidelines for the encoding of dictionary information on written and spoken forms”. In *Electronic Lexicography in the 21st Century: Proceedings of eLex 2017 Conference*, 485-94.
- Bohbot, H., Frontini, Luxardo, F. G., Khemakhem, M., and Romary, L. (2018). Presenting the Nénufar Project: a Diachronic Digital Edition of the Petit

- Larousse Illustré. In *GLOBALEX 2018* – Globalex workshop at LREC2018, May 2018, Miyazaki, Japan, 1-6. <hal-01728328>.
- Bowker, L. 2018. “Lexicography and terminology”. In *The Routledge Handbook of Lexicography*, edited by P. Fuertes-Olivera, P. Abingdon: Routledge Handbooks Online.
- Budin, G., Majewski, S. and Morth, K. 2012. “Creating Lexical Resources in TEI P5: A Schema for Multi-purpose Digital Dictionaries”. In *Journal of the Text Encoding Initiative* 3, <http://jtei.revues.org/522>.
- Costa, R. 2013. “Terminology and specialised lexicography: two complementary domains”. In *Lexicographica*, 29, 29-42.
- ISO 1087-1. 2000. Terminology work – Vocabulary – Part 1: Theory and application. Geneva: International Organization for Standardization.
- Morris, D. 1985. *A Tribo do Futebol*. Lisboa: Publicações Europa-América.
- Romary, L. 2015. “TEI and LMF crosswalks”. In *Digital Humanities: Wissenschaft vom Verstehen*. Berlin: Humboldt Universität zu Berlin, <hal-00762664v2>.
- Romary, L., Tasovac, T. 2018. “TEI Lex-0: A Target Format for TEI-Encoded Dictionaries and Lexical Resources”. In *Proceedings of the 8th Conference of Japanese Association for Digital Humanities*, 274-275, https://tei2018.dhii.asia/AbstractsBook_TEI_0907.pdf.
- Salgado, A., and Costa, R. 2019. Marcas temáticas en los diccionarios académicos ibéricos: estudio comparativo. *RILEX. Revista Sobre Investigaciones Léxicas*, 2(2), 37-63. <https://doi.org/10.17561/rilex.v2.n2.2>.
- Salgado, A. Costa, R. and Tasovac, T. 2019a. Improving the consistency of usage labelling in dictionaries with TEI Lex-0. In *Lexicography: Journal of ASIALEX* 6 (2), 133-156. DOI: <https://doi.org/10.1007/s40607-019-00061-x>.
- Salgado, A., Costa, R., Tasovac, T. and Simões, A. 2019b. TEI Lex-0 In Action: Improving the Encoding of the Dictionary of the Academia das Ciências de Lisboa. In I. Kosem *et al.* (eds.), *Electronic lexicography in the 21st century. Proceedings of the Electronic lexicography in the 21st century (eLex 2019): Smart lexicography*, 417-433), 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o. Retrieved from: https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_23.pdf.
- Silva, R. and Costa, R. 2019. Accéder aux connaissances des experts par l’entremise de la médiation en Terminologie. In M. De Gioia e M. Marcon, eds., *L'essentiel de la médiation. Le regard des sciences humaines et sociales*. Peter Lang, 105-121. DOI: <https://doi.org/10.3726/b16164>. ISBN 978-2-8076-1088-0.

Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Niemeyer.

Dictionaries

Dicionário da Língua Portuguesa Contemporânea, 2001, João Malaca Casteleiro (coord.), 2 vols. Lisboa: Academia das Ciências de Lisboa & Editorial Verbo.

Diccionario de la Lengua Española (24.^a ed.). Real Academia Española [em linha], 2001-2020, www.rae.es/rae.

Dictionnaire de l'Académie Française (9.^a ed.). Académie Française, 2020, <http://www.dictionnaire-academie.fr/>.

Résumé

La complémentarité entre lexicographie et terminologie est le point de départ de cet article. Travaillant dans le domaine du football, cet article examine les termes du football dans trois dictionnaires de l'Académie de Sciences — portugais, français et espagnol — et présente l'encodage TEI du terme « defesa » (défense), qui désigne une position occupée par les joueurs sur le terrain. Après avoir identifié toutes les positions que les joueurs de football peuvent occuper sur le terrain, nous avons vérifié si les dictionnaires incluent ces termes. Une étude contrastive du corpus nous a permis d'observer les données et de tirer des conclusions sur le travail lexicographique réalisé dans les trois langues. En gardant à l'esprit des concepts tels que la réutilisabilité et l'interopérabilité, nous présentons : 1) une comparaison des termes dans les trois dictionnaires en analyse ; 2) le codage du dictionnaire TEI Lex-0, une norme *de facto* pour faciliter l'interopérabilité ; 3) une modélisation TEI cohérente et une description des éléments microstructuraux des entrées lexicographiques.