ORIGINAL ARTICLE

ANACpédia: corpus, terminology case files and sub-areas

ANACpédia: corpus, arquivos de casos de terminologia e sub-áreas

Adriana Ceschin Rieche[1], Fernanda Alves e Silva[2], Marcela Neves de Medeiros[3], Maria Elisabeth Farah Correia Norões[4]

[1] Department of Airspace Control (DECEA, Rio de Janeiro) <adrianaacr@decea.gov.br>
[2] National Civil Aviation Agency (ANAC, Rio de Janeiro) <fernanda.silva@anac.gov.br>
[3] National Civil Aviation Agency (ANAC, Rio de Janeiro) <marcela.medeiros@anac.gov.br>
[4] Department of Airspace Control (DECEA, Rio de Janeiro) <elizabethmefn@decea.gov.br>

ABSTRACT

This paper presents the Online Dictionaries ANACpédia developed by ANAC – National Civil Aviation Agency, with the cooperation of DECEA – Department of Airspace Control, of the Air Force Command. The theoretical approach is based on the principles of Corpus Linguistics and Socioterminology. Examples of terminology case files developed by the team will be provided, as well as the manual term selection methodology, at first, and the subsequent automatic data search from an English, Spanish and Portuguese corpus consisting of written texts published by aviation authorities in these languages. The usefulness of automated search and concordance tools to enrich the dictionaries and the latest research on sub-areas will be discussed.

KEYWORDS: Aviation terminology; Online dictionaries; Corpus; Sub-areas

RESUMO

Este artigo apresenta os Dicionários Eletrônicos Online ANACpédia, desenvolvidos pela ANAC – Agência Nacional de Aviação Civil, contando com a cooperação do DECEA – Departamento de Controle do Espaço Aéreo, do Comando da Aeronáutica. Os princípios teóricos são os da Linguística de Corpus e os da Socioterminologia. Serão apresentados exemplos de dossiês terminológicos desenvolvidos, bem como a metodologia baseada na seleção de termos de forma manual, num primeiro momento, e a posterior pesquisa com busca automática de dados, a partir de um corpus nos idiomas inglês, espanhol e português, composto por textos publicados pelas autoridades de aviação. Discutiremos a utilidade de softwares de pesquisa automática e de concordância para o enriquecimento dos Dicionários e as mais recentes pesquisas da equipe sobre sub-áreas.

PALAVRAS-CHAVE: Terminologia da aviação; Dicionários eletrônicos; Corpus; Sub-áreas

1 Objectives of the paper

This paper has the primary objective of updating the reader on the evolution of terminology management work developed at http://www2.anac.gov.br/anacpedia/ since 2013. The ANACpédia system consists of Online Electronic Dictionaries[11], in Portuguese, English and Spanish, with about eight thousand words, developed by the National Civil Aviation Agency, the Brazilian Civil Aviation Authority, relying more recently on the cooperation of DECEA (Department of Airspace Control), Aeronautical Authority responsible for controlling the Brazilian airspace. The following topics will be addressed: ANAC-DECEA partnership, theoretical principles, current project objectives, terminology case files, and automatic corpus research, sub-areas, examples and perspectives for future work.

2 Background of the ANAC-DECEA partnership

Since September 2013[2], when it was made available on the Internet, ANACpédia began to reach a larger and more varied public. In November 2013, DECEA, represented by two translators with experience in the aviation area, contacted ANAC, showing interest in the work developed by the Agency.

The professional goals of DECEA team and the need for terminological works in the aviation area coincided with the goals of ANACpédia. A partnership was established for the improvement and development of terminology products "facilitating information interchange, and in the integration of language resources for a knowledge-based society" (2) and at the service of society as a whole.

3 Theoretical grounds

In the early ANACpédia work, in 2004, research was intuitively based on theoretical precepts of Corpus Linguistics (CL) and Socioterminology. With the professional maturity of the team, the LC has become part of everyday activities, both as a methodology and as a theoretical approach, since according Finatto (2), LC includes the concept of language harmonized with the communicative-textual perspective of Textual Terminology. In practice, LC uses corpora of a specific area as a basis for development of reference materials such as monolingual or multilingual dictionaries, glossaries, terminology databases, etc., which characterizes the work within ANACpédia. According to Sardinha (3), "[...] Corpus Linguistics deals with the collection and exploration of corpora, or sets of textual linguistic data were carefully collected with purpose to serve for the research of a language or linguistic variety. As such, it is dedicated to the exploration of language through empirical evidence, extracted by computer resources."

The ANAC and DECEA professionals direct their activities based on these theoretical perspectives in an attempt to approach terminological harmonization, conceptual precision, linguistic correctness and appropriateness of the term, aiming at effective communication.

4 Current project objectives

Relying on a number of terms and acronyms, one of the project goals at the moment is to enrich the information through the inclusion of more synonyms, related terms, hypernyms, hyponyms, aviation sub-areas and contexts of use and establish semantic relations, when possible. That is, priority is given to establishing relations between the terminology records, selection of textual support and use documentation in the specialized discourse. According to Antón and Nistal (4), the classification of the entries – which, in the scope ANACpédia, is materialized through the inclusion of aviation sub-areas – considerably affects the use of dictionaries as well as the availability of contextual data (inclusion of "contexts of use" in our work).

ANACpédia seeks to work also in an attempt to arrive at terminological harmonization because it is considered important to have tools available in Portuguese, prepared by technicians having Portuguese as their mother tongue, and which are effectively useful for Portuguese language speakers. We agree with Santos (5) in that "(...) we need to devise tools, advertising, large systems for the Portuguese-speaking public, built by Portuguese researchers and technicians", highlighting "Portuguese-speaking public" and "researchers and technicians whose native language is Portuguese."

The importance of the work explained in the previous paragraph is stressed if we consider that currently the sources of terminological references in the aviation area, in Portuguese, are scarce. Existing sources are often restricted to small groups and are not, as a rule, prepared based on appropriate terminology and linguistic studies.

---

[1] The development of digital and electronic terminology resources has been widely considered due to the obvious ease of search and access to a more extensive public. According to L'Homme & Cormier (8), "*Digital or electronic lexicography (...) has changed dictionaries and lexicography in a profound way and we will probably witness many more changes in the coming years.*"

[2] ANACpédia was already available on the ANAC Intranet since June 2012 and, therefore, had been used by the Agency's servers.

In the 1990s, Santos stated that the Portuguese language research community was not democratic at all in making research results and tools available that could be quite useful to a certain audience (5).

Being available online to the entire Brazilian and foreign companies, ANACpédia may be considered a facilitating instrument to users, having free and unrestricted access, allowing the public to view corpus data consolidated by the team.

5 Preparation of terminology case files

During the work, the team prepares terminology case files that are discussed and validated in meetings,

originating terminology records. In practice, we can define, for our use that a terminological case file is an editable document saved in Word format, containing columns representing the fields to be included in Dictionaries, that is, the case file should represent the essential information we want to make available to users.

The terminology case files consist of information used by aviation professionals in their daily lives. The team interacts with experts in forums, meetings, working groups, workshops, symposia, study groups and training events in general, as well as during daily professional activities. Such interaction enables the assurance that the language decisions are made based on the information we have acquired in "expert environments", through oral and written speech.

Figure 1. Example of terminology case file prepared by the team – new terms suggested

LISTA 10
Sugestões de Novos Termos para ANACpédia – Acesso em novembro de 2014

Termo	Tradução	Definição	Fonte	Definição Port.	Fonte Port.	Notas Adicionais	Notas Ling	Sinônimos	Obs.
altostratus	altostratus	A grayish or bluish cloud sheet or layer of striated, fibrous or uniform appearance; totally or partially covering the sky; and having parts thin enough to reveal the sun at least vaguely.	KUMAR	Nuvens que se apresentam como véu fibroso ou estriado, de cor mais ou menos acinzentada ou azulada. Sua altitude varia entre 2000 e 3000 metros.	ANTAS	Uso (inglês): (*) ICAO ANEXO 3	NA1: Altrostratus does not show halo phenomena.		Sigla: AS ou As
		A uniform layer cloud at moderate altitude.	CROCKER						
cloud base	base de nuvem	The lower surface of a cloud.	KUMAR	Zona mais baixa de uma nuvem na qual o obscurecimento evolui de maneira perceptível, a partir do que corresponde a uma atmosfera clara ou de névoa-seca, para chegar no que corresponde à presença de gotas de água ou cristais de gelo.	ANTAS	Uso inglês: At aerodromes with runways intended for Category I instrument approach and landing operations, automated equipment for measuring or assessing, as appropriate, and for monitoring and remote indicating of surface wind,	NA1: The cloud base is measured from the ground.		Sigla: BASE

Source: Authors.

Figure 2. Example of terminology case file prepared by the team – data supplementing already existing entries

Complementar termos já existentes na ANACpédia:
TERMOS

Termo	Tradução	Definição	Fonte	Definição Port.	Fonte Port.	Notas Adicionais	Notas Ling	Sinônimos	Obs.
line of sight	linha de visada	sugestão: manter apenas a definição 2 que consta na ANACpédia		1 - Linha que une o observador ao objeto ou referência. 2 - Característica de transmissão e recepção dos sinais de rádio, típica das radiofrequências empregadas comumente nas telecomunicações aeronáuticas e nos radares de vigilância de tráfego aéreo. Significa que a transmissão ou recepção se dá em linha reta e não consegue ultrapassar eventuais obstáculos, tal como ocorre com a luz ou a visão humana. Portanto, edifícios, morros ou a própria curvatura da terra podem impor limites de alcance ou até mesmo impedir a comunicação ou detecção por meio de radar.	1 – ANTAS 2 – SIRIUS (http://www.decea.gov.br/sirius/linha-de-visada-2/)	(uso em português) A cobertura da Estação DF, normalmente, não deverá exceder a linha de visada e, consequentemente, será dependente da potência de saída do equipamento de comunicações, da altura da antena, do terreno circunvizinho e dos efeitos de reflexão dos sinais. (CIRCEA 121-3)			Sigla: verificar LOS na lista de siglas
arrivals	setor de desembarque	The part of an airport that deals with passengers who are arriving.	CROKER		Fazer ref. De uma das decisões da junta recursal da ANAC (internet).	Local para desembarque de passageiros num aeroporto.			
arrival	chegada	The act of reaching	CROKER		DUTRA	Exemplos: Gulf			ARR –

Source: Authors.

6 Corpus and automatic research

The types of texts and publications used as research sources and that constitute our corpus are quite varied: ICAO (International Civil Aviation Organization) Annexes; official letters; aeronautical charts; circulars; dictionaries; decrees; guidelines; glossaries; civil aviation instructions; instructions from the Air Force Command; regulatory and supplementary instructions issued by ANAC; legislation; books; aircraft manuals; procedure manuals; works published by various associations; ordinances; publications from the FAA (Federal Aviation Administration – USA); regulations; reports; resolutions; journals published by aviation authorities and ICAO; etc.

Recently, we sought to develop the works based on an organized and structured corpus and automatic data extraction systems and software that enable the visualization and analysis of terms in their occurrence contexts, enabling semantic, grammatical, and lexical analyses. Ultimately, the intention is to enrich the ANACpédia dictionaries establishing semantic relations between the terms, for example.

To this end, the first initiative was to divide our corpus by language. Then, the publications available in electronic format were saved in PDF and TXT, which involved a huge effort considering that the general corpus of ANACpédia currently consists of about 747 reference sources and 10,035,705 words in the three languages.

AntConc (Concordance) freeware (6), designed by Anthony Lawrence, was used for the development of new ANACpédia products and for the improvement of the available dictionaries, to ratify language impressions of the team, provide examples of usage as well as supportive data for the definition of sub-areas, survey of co-occurrences and frequencies in a real corpus.

7 Sub-areas

Sub-areas can be briefly defined as specialty areas hierarchically subordinate to a larger area. Sub-areas are used in ANACpédia to classify terms and acronyms according to their occurrence in the wide universe of our corpus.

Figure 3. Example of a list of sub-areas prepared by the team

Aerodesporto Air Sports É toda atividade não-comercial voltada para a prática do esporte, do turismo e do lazer em que se utilizam engenhos aéreos e esteja prevista no Código Desportivo da Federação Aeronáutica Internacional – FAI. AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. RBAC 01: Regulamentos Brasileiros de Aviação Civil. Definições, regras de redação e unidades de medida. Brasília, 2008. Disponível em: http://pergamum.anac.gov.br/arquivos/RBAC01EMD01.PDF. Acesso em: 17 fev. 2011.	Deportes Aeronáuticos Sport Aérien
Aerodinâmica Aerodynamics (1) Parte da física que estuda as leis reguladoras do movimento dos fluídos elásticos e da pressão do ar externo; estudo do movimento dos fluídos gasosos e da ação de sua força sobre os sólidos em movimento. (2) Estudo das propriedades e características do ar. CAMPOS, M. C.; PAIVA, P. T.; PRICEVICIUS, T.; VARANDA, T. V. C. Glossário de aviação. São Paulo: Núcleo de Estudos de Tradução, 2002.	Aerodinámica Aérodynamique
Aeronautas Crew Aeronauta é o aeronavegante que exerce função remunerada a bordo de aeronave civil nacional. É também aeronauta aquele que exerce função em aeronave civil estrangeira, mediante contrato de trabalho, regido por leis brasileiras. BRASIL. Comando da Aeronáutica. Departamento de Aviação Civil. RBHA 67: inspeção de saúde e certificado de capacidade física. Rio de Janeiro, 1999. 39 p. (Regulamento Brasileiro de Homologação Aeronáutica)	Aeronautas Personnel Navigant Technique

Source: Authors.

Sub-areas were not always present in the ANACpédia dictionaries. Although there was an internal classification from the beginning, until 2014, ANACpédia did not inform the users in which sub-areas the entries were included. The decision to improve and disclose such a classification was taken in view of the future perspective of making the Dictionaries available in "domain tree format"[3]. The team found that, although this is a very complex task, it was necessary to start the classification consolidation to help the team decisions about synonyms and selected contexts of use, for example. A term or acronym may have different meanings according to the sub-areas where they may occur. The decision was supported by the statement by Bononno (7) that "technical dictionaries – at least the better ones – often provide labels to indicate the field of application or domain for a headword".

For the best development of our research, we have prepared a list of sub-areas and a definition for each of them, as exemplified in Fig. 3, above. We understand that a defining text for each sub-area, prepared by us from consultations with dictionaries and other references, greatly guides the work of categorization of terms and acronyms.

8 Examples

Since the first use of AntConc software (Concordance) (6), it was possible to identify the usefulness of the results with immediate display of combinations of terms, contexts of use and sub-areas, according to the matters dealt with by the documents in which the terms occur (Airport Infrastructure, Air Navigation, Operations, Safety, etc.).

The terminology records of ANACpédia concerning the terms aerodrome/*aeródromo* (equivalent in Portuguese), widely used in aviation, are clear but succinct. Research with extraction software and concordance relations can enrich them. In addition, they may promote the establishment of semantic relationships between related terms in English and Portuguese, such as aerodrome control tower; aerodrome traffic; aerodrome climatological summary; aerodrome identification sign; aerodrome operating minima; aerodrome taxi circuit; *aeródromo civil; aeródromo militar; aeródromo de partida; aeródromo de destino; aeródromo de alternativa de destino; aeródromo controlado*; among others.

Figure 4. Term available before research with extraction software (Authors)

Inglês/Português

4 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

<< aeroclub | aerodrome | aerodrome beacon >>

Back to: "A"

aerodrome

Definição1
    A defined area on land or water (including any buildings, installations and equipment) intended to be used either wholly or in part for the arrival, departure and surface movement of aircraft.
Fonte1
    INTERNATIONAL CIVIL AVIATION ORGANIZATION. International civil aviation vocabulary. 3rd ed. Montreal, 2007. (Doc. 9713)
Nota adicional1
    Acronym in English: AD.
Nota adicional2
    Information regarding this term is a result of researches developed by DECEA in cooperation with ANAC.
Subárea
    Airport Infrastructure
Português
    aeródromo

Search in Google | Search in Yahoo | Feedback

---

[3] A domain tree is a scheme containing vocabulary of a particular area, and that represents sub-areas, helping one to understand and organize a field of knowledge.

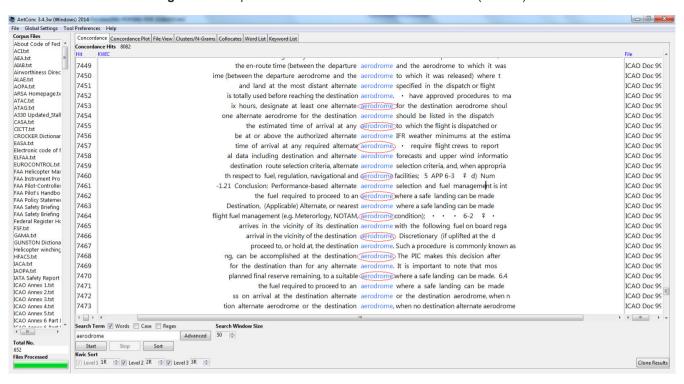**Figure 5.** Example of research in AntConc for related terms (Authors)

Figure 6. Term available before research with automatic extraction software (Authors)

Figure 7. Example of research in AntConc for related terms (Authors)

9 Future work

ANACpédia allows, by its nature, almost endless possibilities for actions, activities, works, development of new Dictionaries and databases, ontologies, classification systems, etc., not to mention academic studies on the Project.

Therefore, given the numerous possibilities, it is important that the team plan their actions in order to set priorities, while considering actions in the medium and long term.

At the time of writing of this article, there is consensus that we need to continue to establish conceptual relations between different records; select the necessary textual evidence to describe the concepts; and document the use of terms in specialized discourse, with the aid of phraseology units. These actions are already represented in the Dictionaries, albeit on a small scale.

Apart from improvements in Dictionaries, the team intends to soon make available a monolingual dictionary in Portuguese, a database comprised of aviation acronyms, also exclusively in Portuguese, and a bilingual Dictionary English-Spanish/Spanish-English.

References

1. Pavel S, Nolet D. Manual de terminologia [Internet]. [cited 2015 January 12]. Available from: <http://linguisticadocumentaria.files.wordpress.com/2011/03/pavel-terminologia.pdf>.

2. Finatto MJB. Exploração terminológica com apoio informatizado: diálogos entre terminologia e linguística de corpus. In: Lorente, M.; R. Estopá, J. Freixa, J. Martí, C. Tebé (ed.) Estudis de linguistica y de linguistica aplicada en honor de M. Teresa Caré Castellví. IULA. 2007; 2.

3. Sardinha TB. Linguística de Corpus: histórico e problemática. São Paulo: Delta; 2000; 16(2).

4. Antón MTO, Nistal, PF. Expert knowledge representation in bilingual e-dictionaries: a case study. Cad. Trad. 2015; 35(1).

5. Santos D. Processamento computacional da língua portuguesa: documento de trabalho [Internet]. [cited 2014 December 29]. Available from: <http://www.linguateca.pt/branco/proc_comp_port.html>.

6. Anthony, L. AntConc (3.4.3w Windows version) [Computer Software] Tokyo, Japan: Waseda University; 2014. Available from: <http://www.antlab.sci.waseda.ac.jp/>.

7. Bononno, R. Terminology for translators: an implementation of ISO 12620. Meta, XLV, 4, 2000.

8. L'homme, MC; Cormier, MC. Dictionaries and the digital revolution: a focus on users and lexical databases. International Journal of Lexicography. 2014; 27(4). http://dx.doi.org/10.1093/ijl/ecu023