http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

Integrating Terminological Resources in Dictionary Portals: The Case of the *Diccionarios Valladolid-UVa*

Pedro A. Fuertes-Olivera, *Department of Afrikaans and Dutch, University of Stellenbosch, South Africa and International Centre for Lexicography, Universidad de Valladolid, Valladolid, Spain (pedro@emp.uva.es)*
and
M.A. Esandi-Baztan, *Department of Materials Science and Metallurgical Engineering, Graphic Expression in Engineering, Cartographic Engineering, Geodesy and Photogrammetry, Mechanical Engineering and Manufacturing Process Engineering, Universidad de Valladolid, Spain (mariaangeles.esandi@uva.es)*

Abstract: This paper advocates the convergence of terminology and lexicography, and illustrates this view by presenting some of the steps taken for incorporating terminological resources and ideas in an online dictionary portal that is being constructed at the University of Valladolid (Spain). This dictionary portal contains several dictionary types, was designed by the same team and is being constructed from the same theoretical perspective, regardless of whether some of the lexical items included are judged "lexicographic", i.e. related to *general language expressions*, or "terminological", i.e. connected with *terms*. In addition to dealing with certain basic tenets of dictionary portals, the paper describes an ad-hoc typology of *definitions* that has been created for two main reasons. Firstly, it makes the process of compilation easier, more uniform, and more readily systematised, thus facilitating the efforts of different people in different places at different times. Secondly, these definitions will feed the Spanish–English *Write Assistant*, a commercially driven language tool that uses a language module based on statistics and is in the process of using Artificial Intelligence (AI) technologies, e.g. machine learning and neural networks, for creating patterns. We have found that precise definitions, similar to terminological (i.e. encyclopaedic) definitions, for most lemmas increase the tool's functions. Such definitions offer a very different picture of current monolingual Spanish and bilingual Spanish–English dictionaries.

Keywords: DICTIONARY PORTAL, ONLINE DICTIONARIES, DEFINITION, ENGLISH, SPANISH, CONVERGENCE OF TERMINOLOGY AND LEXICOGRAPHY, AI TECHNOLOGIES

Opsomming: Die integrasie van terminologiese hulpbronne in woordeboekportale: Die geval van die *Diccionarios Valladolid-UVa*. In hierdie artikel word die konvergensie van terminologie en leksikografie bepleit, en hierdie siening word geïllustreer deur sommige van die stappe vir die inkorporering van terminologiese hulpbronne en idees in 'n aanlyn woordeboekportaal wat by die Universiteit van Vallodolid (Spanje) saamgestel word,

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

voor te lê. Hierdie woordeboekportaal bevat verskeie woordeboektipes, is deur dieselfde span ontwerp en word vanuit dieselfde teoretiese perspektief saamgestel, ongeag of sommige van die leksikale items wat ingesluit word as "leksikografies" beskou word, m.a.w. verwant aan *algemene taaluitdrukkings*, of as "terminologies", m.a.w. verwant aan *terme.* Benewens die hantering van sekere basiese woordeboekportaalbeginsels, beskryf die artikel ook 'n ad-hoc-tipologie van *definisies* wat hoofsaaklik om twee redes geskep is. Eerstens vergemaklik dit die samestellingsproses en maak dit ook eenvormiger en makliker om te sistematiseer. Sodoende word die pogings van verskillende mense op verskillende plekke en tye gefasiliteer. Tweedens sal hierdie definisies as bron dien vir die Spaans–Engelse *Write Assistant*, 'n kommersieelgedrewe taalhulpmiddel wat gebruik maak van 'n taalmodule wat op statistiek berus en wat Kunsmatige Intelligensie (KI)-tegnologie, bv. masjienleer en neurale netwerke, inspan vir die skep van patrone. Ons het bevind dat presiese definisies, soortgelyk aan terminologiese (m.a.w. ensiklopediese) definisies, vir die meeste lemmas die funksies van die hulpmiddels uitbrei. Sulke definisies bied 'n heeltemal ander blik op bestaande eentalig Spaanse en tweetalig Spaans–Engelse woordeboeke.

Sleutelwoorde: WOORDEBOEKPORTAAL, AANLYN WOORDEBOEKE, DEFINISIE, ENGELS, SPAANS, KONVERGENSIE VAN TERMINOLOGIE EN LEKSIKOGRAFIE, KI-TEGNOLOGIE

1. Introduction

Terminology may be defined as "the study of and field of activity concerned with the collection, description, processing and presentation of terms, i.e. lexical items belonging to specialized areas of usage of one or more languages" (Sager: 1990: 2). This view connects terminology with lexicography, a process that the coming of age of the Internet, the preponderance of the user paradigm, and the influence of corpus methodology has accelerated. Bowker (2018: 138) summarises this approach, which has been well attested since the 1990s, by commenting on Sager's suggestion that "in its objectives, terminology is akin to lexicography which combines the double aim of collecting information about the lexicon of a language with providing an information — and sometimes even an advisory — service to language users." She (2018: 147) documents her stance with existing practices in both camps, in which we can observe that dictionaries are incorporating terms and that terminological resources are adding general language expressions:

> Both lexicographers and terminologists have indeed taken up this challenge to produce resources that meet user needs. For instance, in response to user requests, the second edition of the *Macmillan English Dictionary* (2007) now incorporates a range of specialist terms along with its general language offering. Meanwhile, term bases for private corporations, as well as those developed for use by translators, now regularly include general language expressions alongside specialized terms.

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

Bowker (2018: 147) concludes her analysis on the relationship between both disciplines by claiming that "although some differences remain, the two disciplines nonetheless appear to be converging with regard to many aspects of their practice." Table 1 shows how she views the evolution and convergence of characteristics of lexicography and terminology:

	Lexicography	**Terminology**
Practitioner	mainly lexicographers, but with greater involvement from the general public (via crowdsourcing)	mainly terminologists, but with greater involvement from the general public and subject matter experts (via open and closed crowdsourcing)
Object of study	mainly words, but also some terms	mainly terms, but also some general language words or expressions
Domain	mainly general language, but also some specialised language	mainly the language of a specialised domain, but also some general language
Point of view	mainly descriptive	mainly normative/prescriptive in the public and academic sectors, but incorporating more descriptive elements in commercial settings
Approach	mainly semasiological	increasingly semasiological, but retaining some onomasiological elements where useful
Organisation	mainly alphabetical, but sometimes incorporating thematic elements	mainly thematic, but allowing alphabetic searching
Main information provided	words, meanings, examples, usage information (e.g. collocations, frequency, phraseology), a range of linguistic information (e.g. part of speech, pronunciation)	preferred term, variants, context and usage information (e.g. collocations, frequency, phraseology), meaning, conceptual relations
Intended users	lay people, professional and academic audiences	public sector (for language planning), domain experts, scientific/technical writers, translators (for bi- or multilingual resources), commercial enterprises

Table 1: Evolution and convergence of characteristics associated with lexicography and terminology. Source: Bowker (2018: 148)

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

This paper elaborates on Bowker's main conclusions and adds several reasons for advocating the integration of lexicography and terminology. Firstly, instead of single dictionaries, we have to design and construct dictionary portals, where traditionally associated differences between lexicography and terminology can be downplayed or even eliminated (Section 2). Secondly, integration will work better if the dictionary portal contains *up-market online resources*, as these will be based on the use of *disruptive technologies* that are necessary for designing and constructing commercially-driven language tools. Thirdly, the nature of these tools is not affected by the data type they contain — i.e. whether we work with *words* or *terms* — but by the philosophy underlying them and the technical and technological resources used for dealing with their computational and linguistic aspects (Section 3). Finally, the integration process is illustrated with an ad-hoc typology of *definitions*, i.e. a collection of different types that was created for several reasons. For instance, this typology makes the process of compilation easier, more uniform, and more readily systematised, thus facilitating the efforts of different people in different places and at different times (Section 4). A final conclusion will summarise our main findings and will enumerate certain topics that may merit more attention in future research.

2. Dictionary portals

Engelberg and Müller-Spitzer (2013: 1023) define a dictionary portal as "a data structure that is presented as a page or set of interlinked pages on a computer screen and provides access to a set of electronic dictionaries, and where these dictionaries can also be consulted as standalone products". With the underpinning of several criteria — type of access provided, cross-referencing structures, ownership relations between the portal and the dictionaries and layout of the portal — Engelberg and Müller-Spitzer (2013) propose a typology of dictionary portals comprising (a) dictionary nets, (b) dictionary search engines, and (c) dictionary collections. Their analysis, as well as that of Boelhouwer et al. (2018), concludes that dictionary portals are widespread and that they may illustrate the way ahead for future online dictionaries.

The above definition merits some comments. Firstly, a dictionary portal is a "data structure". This can offer users more than existing corpus-based online dictionaries provided that designers of such portals assume that data structures might be analysed with big data analytics. For instance, designers of the project *Diccionarios Valladolid-UVA* (see Section 3, below) have used big data analytics for analysing around 60 million look-ups in existing online dictionaries; the aim was to discover real searches, for which an initial lemma list of 20,000 English words and 16,000 Spanish words was created. This process has comprised several stages and is based on an analysis of around one million daily searches in several dictionaries, e.g. an English–Spanish/Spanish–English dictionary, an English–German/German–English dictionary, an English monolingual dictionary, a Spanish monolingual dictionary, and so on. Around 80% of the searches

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

can be matched, i.e. the same search is identified in the logfiles of different dictionaries and can, therefore, be interpreted as an identification of the most popular dictionary articles in the dictionaries under scrutiny. After two months of work with the search logfiles, which amounted to more than 60 million look ups, IT professionals working on this project were able to produce the above-mentioned lemma lists (Fuertes-Olivera 2019).

Secondly, the data structure will be "presented as a page or set of interlinked pages on a computer screen." It is obvious that mobiles and similar devices, e.g. tablets, are also carriers of data structures and that most people view them on these screens rather than on those of computers. This is influencing the design of online dictionaries, which must make provisions for allowing users to consult data on smaller screens than those originally conceived.

Thirdly, it "provides access to a set of electronic dictionaries". This sentence is clearly confusing, as most existing dictionary portals only contain retro-digitised dictionaries, i.e. printed dictionaries converted to a digital format. An analysis of the 37 dictionary portals identified in Boelhouwer et al. (2018: 765-767) has allowed us to classify them into three main types, the criterion being the presence of real online dictionaries.

The *Free Dictionary.com* illustrates the first type. It *mostly* contains data from printed reference works — typically from dictionaries, glossaries or encyclopaedias — and from free access online ones, e.g. *Wikipedia.* For instance, the entry for *cost accounting* in *The Free Dictionary.com* offers the following data:

— Definitions from the *Collins English Dictionary* 12th edition, and the *Random House Kernerman Webster's College Dictionary.*
— The symbol "n" for noun.
— The related words *cost accountant* with the symbol "n" for "noun" and a definition extracted from the *American Heritage® Dictionary of the English Language, fifth edition*, as well as *accountancy, accounting* and *costing*.
— An image that connects *cost accounting* with *costing, accountancy,* and *accounting*, without explaining the connection.
— A very long encyclopaedic article of *accounting* from the *West's Encyclopedia of American Law, 2nd edition*, published in 2008. In this long entry we are told that various accounting methods are employed and that one of them is the *cost method of accounting*. This article includes a section on "Further readings" and cross-references to *Accrual Basis; Cash Basis*; and *Income Tax*.
— A Definition from the *Farlex Financial Dictionary*.
— A Definition from the glossary *Wall Street Words: An A to Z Guide to Investment Terms for Today's Investor*, authored by David L. Scott.
— An encyclopaedic article from the *Great Soviet Encyclopedia, 3rd Edition*, published between 1970 and 1979.
— An encyclopaedic article from *Wikipedia.*

Users searching for *cost accounting* in *The Free Dictionary.com* will face many

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

problems, three of which are relevant for this paper: (a) they will not be sure about the meaning and usage of *cost accounting* in today's financial world, as they will come across contradictory information, e.g. *cost accounting* is defined in the context of the Soviet Union of 1970, Wall Street and the United States of 2008 (before the Great Recession and after the Enron Failure), and the changes introduced after the Great Recession so as to make future financial crises more difficult; (b) the consultation process takes a lot of time and energy, as users must decide for themselves which, if any, of the examples of contradictory information they find suitable; and (c) there are no usage contexts, e.g. grammatical information, example sentences, and so on.

Lexico.com is an example of the second type (see at https://www.lexico.com/). It includes free and restricted areas, the latter for subscribers, and especially for users of *English* and *Spanish dictionaries.* In this portal, users access printed and online information tools. For reasons of space and convenience, we will only analyse the free part of the portal, in which we can look up in English and Spanish online dictionaries as well as a Thesaurus, an English grammar and a Spanish grammar. The dictionaries included in the portal contain features of both online dictionaries, e.g. they do not use abbreviations, and printed dictionaries, e.g. they continue using *recursive definitions* and definitions by synonyms. Example 1 shows the definition of *accrual accounting* in the English Dictionary. This is an example of *recursive definition,* i.e., users must search "accrual basis" and "accounting" in order to understand its meaning:

accrual accounting
The practice of accounting on an accrual basis.

Example 1: Definitions of *accrual accounting* in *Lexico.com*

English is the node language in this dictionary portal and it seems to have considerable influence on how dictionaries are designed, compiled and updated. Having a node language makes design and construction processes cheaper, as data can easily be made reusable, i.e. transferred from, say, a monolingual English dictionary to a bilingual English–Spanish one. The grammar tabs display grammatical information, e.g. definitions of grammatical concepts such as "personal pronoun" and so on. The thesaurus offers synonyms and antonyms under an example sentence.

Searching for *cost accounting* in the free area has allowed us to extract the following information:

— A definition.
— Part of speech (NOUN).
— Grammar information: it is a mass noun.
— Several example sentences, all of which seem to have been extracted from a specialised corpus.

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

— Pronunciation. Users can hear the pronunciation of the word and read its phonetics.
— Equivalents in Spanish and, for some lemmas, translations of example sentences in Spanish. For instance, we have the Spanish equivalents *contabilidad analítica de costes* and *contabilidad analítica de costos* (the former is used in European Spanish and the latter in American Spanish; although important, this information is missing).

Looking up this dictionary portal, users will also face several difficulties, three of which are no doubt problematic for Spanish users:

— the definition of concepts is poor: it lacks reference to concepts, domains and the like. For instance, we are not informed that *cost accounting* is an accounting method;
— users have to infer the context of usage, as the grammar given is rather weak (for instance, the English–Spanish part does not inform that *cost accounting* is uncountable) and usage expressions are scarce or non-existing for some meanings and usages. By way of example, *accounting board* has two meanings in the *English Dictionary,* but all the English sentences and their Spanish translations are on a different interface and are accompanied by the warning that the portal cannot verify their accuracy; and
— the context sentences work as "proxies", i.e. they can or cannot refer to the concept, which force users to decide for themselves whether or not they are suitable.

OWID is an example of the third type. It is a platform "which integrates multiple dictionaries via a common interface, with a joint lemma list and single search option" (Storjohann 2018: 574). What makes this portal different from the other two types is that it *also* offers external resources so as to integrate different academic lexicographic resources with the focus on contemporary German. In *OWID* there is a kind of macrostructure, i.e. a list of alphabet letters at the top of the page that can be clicked to recover the lemmas initiated with the letter chosen. The lemma list is shown on the left of the page and can be modified by clicking on the up and down arrow-like symbols that accompany the lemma list. Searching *Kostenrechnung* (*cost accounting* in German) retrieves the following data:

— Its orthography, including syllabic divisions, etymology and sub-comments on semantics.
— Its grammar and "context profile", i.e. which words typically go with it. This is achieved by cross-referencing users to *Canoonet,* also a German dictionary portal with precise grammar descriptions of each German word, and to *CCCB,* a lexical database that offers contexts, contrasts, lexicographic profiles, semantic proximity and related collocations of each lemma searched for. In other words, it offers lexical maps, i.e. collocational profiles of lemmas.

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

— Sentence examples taken from the *elexiko corpus.*
— Automatic selection of texts that aim to offer a kind of definition(s) of the lemma.

For the purpose of this paper, *OWID* is interesting as it includes several types of dictionaries with other reference tools. For instance, users can access an online dictionary of contemporary German, a dictionary of Neologism, a collocations dictionary and a "Discourse Dictionary", i.e. a sort of dictionary containing lexical maps. These resources are integrated, i.e. they are not a random collection of unrelated dictionary tools, but a collection of lexical data organised through an innovative concept of data modelling and structuring (Müller-Spitzer and Möhrs 2008). Following this line of work, we also defend integrating lexicography and terminology in the same dictionary portal, as shown below.

3. Integrating terminological resources in up-market dictionary portals: The *Diccionarios Valladolid-UVa*

The *Diccionarios Valladolid-UVa* were initially designed as independent online dictionaries in 2012. However, in 2017 they were re-designed and integrated in a new type of reference tool termed "integrated dictionary portal" and defined as

> A reference tool whose Dictionary Writing System is equipped with disruptive technologies. These allow lexicographers to store as much data as possible and users to retrieve only the data they need in specific use situations. Its articles are prepared by the same team with the basic aim of helping human and/or machine users to meet their needs in a quick and easy way. They contain both lexicographically prepared data and linked open data with lexicographic value. The lexicographic data is reusable, subject to a constant process of updating and can be used in conjunction with other tools, e.g. assistants. (Fuertes-Olivera 2016)

At the time of writing this paper, this dictionary portal is still under construction. For this purpose we are using three Dictionary Writing Systems (DWS), each of which is devoted to one particular dictionary type:

(a) monolingual Spanish dictionaries;
(b) bilingual English–Spanish/Spanish–English dictionaries; and
(c) specialised dictionaries.

Each DWS contains an editor, database, Web interface, various management tools, and a kind of dictionary grammar that specifies the structure of the portal, e.g. by making ad-hoc connections of the different elements integrating the DWS (Kilgarriff 2006: 7). They are *currently* storing around 128, 000 general Spanish lemmas, approximately 22, 000 English lemmas, and some 12, 000 English and Spanish specialised lemmas. Around 15,000 more lemmas are added per year and it is expected to go public in 2022. By that time we expect to have finished the lexicographic description of some 75,000 lemmas and to have pre-

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

pared different access structures and interfaces. We aim to create dynamic access structures, i.e. users will have at their disposal different access possibilities, and easy-to-consult interfaces, e.g. all without publicity. Depending on its possible commercial success, this portal can be increased by more specialised and general dictionaries.

Comparing the *Diccionarios Valladolid-UVa* with the above types of dictionary portals, especially with OWID, there are several important differences; three are relevant for this paper. Firstly, this portal is integrated, which signifies the following:

(a) that all the dictionaries included in the portal use a common grammar, that is, an abstract representation of dictionary structure, e.g. the same access structure (Koeva and Blagoeva 2013);
(b) that much of the lexicographic data is reusable, i.e. it can be transferred from one Dictionary Writing System to another;
(c) that all of them are conceptualised as reference tools (Fuertes-Olivera and Tarp 2014);
(d) that they will also support and feed *Write Assistant* (see section 5 and Fuertes-Olivera and Tarp 2020), and
(e) that they *include all members of any clearly-defined sets*, e.g. accounting methods, days of the week, poisonous snakes, non-poisonous snakes, chemical elements, etc. For instance, there are three lemmas for Spanish *pez* (English fish), and one lemma for including all the components of the "fish taxonomy": *pez rata*, *pez luna*, *pez rubia*, *pez sable*, *pez zorro*, *pez mujer*, *pez negra*, *pez gordo*, *pez sierra*, *pez ángel*, *pez payaso*, *pez obispo*, *pez piloto*, *pez resina*, *pez griega*, *pez espada*, *pez volador*, *pez reverso*, *pez machete, pez rémora*, *pez martillo*, *pez trompeta*, *pez mantequilla*, *pez aguja, pez ballesta, pez clavo, pez cofre, pez de San Pedro, pez limón, pez erizo, pez gato, pez gordo, pez guitarra, pez hacha, pez loro, pez pulmonado, pez rojo de China, pez torpedo, pez vela, estar pez* and *que le folle un pez que la tiene fría.*

The "treatment of *pez* illustrates some of the features of this dictionary portal. *Pez* has three grammars. *Pez* (English *fish*) is an animal that admits singular and plural forms, but *pez* (English *tar*) is also a substance that is used only in the singular and *pez* (English *have no idea)* is an invariable adjective that is typically used in multiword lemmas such as *estar pez*. Other lemmas are included because they refer to things, i.e. concepts. Lemmatising them allows us to include categories and taxonomies, e.g. the Spanish taxonomy *peces* (fish) shown above. All the members of these clearly defined sets are accompanied by their Latin formal names as synonyms and can thus be easily identified in several languages, e.g. in English and in Spanish. Finally, we have also included lemmas referring to facts, e.g. cities of the world, rivers, names of battles, festivities, and so on.

Secondly, each meaning of a lemma can be described with up to 67 different lexicographic data. Examples 2 and 3 show the dictionary article *balance*

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

sheet in two use situations (users will have more than 28 different use situations at their disposal): a general English–Spanish production situation for Spanish native speakers (example 2), and a specialised English production situation for English native speakers and translators (example 3):

balance sheet UK English

Gramática
Nombre contable, usado en singular y plural.
Flexiones
balance sheet, balance sheets
Nota de uso
La forma "balance-sheet" se usa menos (Ngram Viewer).
Definición

> documento financiero que recoge el informe en el que se analiza el activo, o bienes o valores de los que dispone la empresa, y el pasivo, u obligaciones económicas de la misma como deudas, préstamos, etc., y su patrimonio neto, la diferencia entre el activo y el pasivo de la empresa; este análisis permite conocer la situación económica y financiera de una empresa o entidad en un momento determinado; se usa en contabilidad

Sinónimos
balance
balance-sheet
statement of financial position US English

Equivalent
balance de situación

Collocations
a condensed balance sheet
un balance de situación consolidado

account form balance sheet
balance con forma de cuenta

derecognise in the balance sheet
eliminar del balance
(…)

Example
The balance sheet is a statement of the enterprise's assets, equity and liabilities at the balance sheet date.
El balance de situación es un document que registra los activos, pasivos y patrimonio de la empresa en la fecha en la que se formula el balance.

Formations (compuestos y derivaciones)
balance sheet account
cuenta de balance

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

balance sheet at year-ends
fecha de cierre de balance

balance sheet date
fecha de cierre de balance

balance sheet entry
entrada en el balance

balance sheet format
formato de balance

balance sheet liability method
método de la deuda basado en el balance

balance sheet sum
suma del balance

balance sheet total
total del balance

balance sheet value
valor del balance
(…)

Example 2: The entry *balance sheet* in the *Diccionarios Valladolid-UVa*: Production in an English–Spanish *general* situation (English) (excerpts)

balance sheet UK English
noun
Inflections
A balance sheet, the balance sheet, balance sheets
Definition
The balance sheet is a statement of the enterprise's assets, equity and liabilities at the balance sheet date. The statement is a status report estimating the enterprise's assets, equity and liabilities as a snapshot at a certain date.
Synonyms
Balance
Statement of financial position US English
Collocations

- a comparative balance sheet as of the end of the immediately preceding financial year
- balance sheet amounts
- balance sheet at the end of the current interim period
- balance sheet at year-end
- balance-sheet layout
- disclose on the face of the balance sheet
- (…) (twenty more collocations, i.e. unfinished sentence examples)

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

Examples
A vertical balance sheet is one in which the balance sheet presentation format is a single column of numbers, beginning with asset line items, followed by liability line items, and ending with shareholders' equity line items.

Example 3: The entry *balance sheet* in the *Diccionarios Valladolid-UVa*: a specialised English production situation (excerpts)

The above examples 2 and 3 represent only 7.4% of the presentation possibilities available at the time of writing this paper (there will be around 28 different possibilities in the *Diccionarios Valladolid-UVa*). For reasons of space, suffice it to say that each of these is based on technological and lexicographic decisions (Fuertes-Olivera, Tarp and Sepstrup 2018), and together they account for the third important feature of this project. Technological and lexicographic aspects are based on the concept of *disruptive innovation* espoused in business theory and practice. Disruptive innovation was introduced in 1995 and has proved a powerful way of thinking about innovation-driven growth, insofar as disruption "describes a process whereby a smaller company with fewer resources is able to successfully challenge established incumbent businesses" (Christensen et al. 2015).

In our case, we have prepared, designed and compiled up-market reference tools, i.e. tools that aim at displacing established competitors by offering a whole new population of consumers at the base of a market access to a product or service that was only accessible to consumers with a lot of money or a lot of skill (Christensen 2011). For instance, preparing *dynamic dictionary articles,* i.e. different data for different users in different situations (examples 2 and 3 above), is a feature of up-market online dictionaries that can easily be implemented as a strategy for broadening the customer base of online dictionaries. Secondly, we have prepared our dictionaries as part of the "data-driven economy", e.g. we are using many sources and a large amount of data (Fuertes-Olivera 2015). An example of this is the type of definitions used in our general dictionaries. On average, they contain around 75 words per definition and resemble *terminological definitions* (see example 2). They are crafted following an ad-hoc typology that has been engineered with two main aims, namely, to help lexicographers in their daily work and to facilitate the creation of patterns that can be displayed by using AI technologies, especially machine learning and neural networks (see below).

4. Definitions

The concept of definition has been the subject of scrutiny in different fields, e.g. Philosophy, Logic, Law, Linguistics, Terminology and Lexicography. For the purpose of this paper, definitions describe the meaning of the lemma, i.e. the

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

"set of conditions which must be satisfied by a lexical unit in order to denote the extralinguistic reality/-ies which correspond(s) to each of its senses" (Fuertes-Olivera and Arribas-Baño 2008: 69). Hence, they refer to the "specific set of data that explains the meaning of a lemma and which is clearly addressed to the lemma" (Nielsen 2011: 202).

Rundell (2015: 314) indicates that, in the printed era, a focus on economy led to definitions "which achieve conciseness (and aspire to precision) through the use of standard formulae ("the act of X-ing), "characterised by Y", and so on) and through a "recursive" strategy." These strategies have costs which are passed on to the user, who has to learn these conventions in order to understand what the dictionary is saying. He adds that in the last 30 years publishers, and especially those in the UK, have addressed this issue by developing more open defining styles. These aim to offer enough information for understanding the definition without knowledge of *dictionarese*, i.e. the typical dictionary conventions such as *recursiveness*.

After reviewing the literature on definitions, Fuertes-Olivera and Arribas-Baño (2008: 70) report that so-called *lexical definitions*, *conceptual definitions*, *relational definitions*, *definitions by extension*, *definition by intension*, *partitive definitions*, and *encyclopaedic definitions* show only a few formal differences among the defining styles of the specialised dictionaries they study. In the *Diccionarios Valladolid-UVa* we have focused on these differences with the aim of constructing an ad-hoc typology that assumes the tenets of the so-called *integrationist approach* (Harris and Hutton 2007). The basic assumptions here are that all signs are semantically indeterminate, that meanings are lexicographers' or terminologists' creations, and that these rely on precise and specific contexts:

> A lexicographical definition (...) does not identify a meaning independently existing in actual usage and *discovered* there by the lexicographer: it is deliberately *constructed* and *allocated* by the lexicographer on the basis of materials selected for study, and its allocation will depend on the viewpoint the lexicographer has chosen to adopt. (Harris and Hutton 2007: 78)

and

> A definition can only be as effective as the context allows it to be, and the context includes the situation of the person seeking to understand the meaning. The notion of a definition adequate to all occasions and all demands is a semantic *ignis fatuus*. (Harris and Hutton 2007: 49)

Our typology comprises 8 types of definitions:

1. *Specialised definitions*: those which describe the meaning of abstract concepts; basically speaking, ideas and thoughts that are not part of the material world but human constructions used in specialised texts for denoting the basics of a particular domain. We use several full sentences that go from more general to more specific aspects. With these definitions we always *include* example sentences that help one to understand the concept.

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

All such definitions are included in the specialised dictionaries of the portal, crafted by experts in the field, and *mainly* target human users, e.g. experts and experienced translators (example 4):

coste atribuido (English: deemed cost)
Definición
El coste atribuido es el importe usado como subrogado del coste o del coste depreciado en una fecha determinada. En la depreciación o amortización posterior se supone que la entidad había reconocido inicialmente el activo o pasivo en la fecha determinada, y que este coste era igual al coste atribuido.
Equivalente
deemed cost < a deemed cost, the deemed cost, deemed costs>
Ejemplo
- Las partidas de edificios, instalaciones y equipos se valoran al coste como coste atribuido menos amortización acumulada y deterioros.
- Items of property, plant and equipment are measured at cost as deemed cost less accumulated depreciation and impairment losses.

Example 4: Example of *specialised definitions* in the *Diccionarios Valladolid UVa* (Spanish–English Bilingual specialised dictionaries).

2. *Semi-specialised definitions*: those which describe the meaning of abstract concepts that have been subjected to a process of determinologisation and that can, therefore, be *also* found in non-specialised texts. We use two main defining styles, a full-sentence Cobuild-style definition in a specialised dictionary and a gloss in a general dictionary (examples 5 and 6, respectively); they are crafted by lexicographers and basically target human translators and interested laypersons:

cancellation
Definición
Cancellation is the act of bringing an arrangement, e.g. a contract, to an end as from a particular date.

Example 5: Example of *semi-specialised definitions* in the *Diccionarios Valladolid UVa* (English specialised dictionaries).

cancelación (English: cancellation)
Definición
anulación del efecto de una obligación jurídica

Example 6: Example of *semi-specialized definitions* in the *Diccionarios Valladolid UVa* (Spanish general dictionaries).

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

3. *Special definitions:* those which describe the meaning of "entities" present in the physical world, e.g. animals, objects, plants, etc., i.e., someone or something that can be pictured. They contain chunks of clauses that are juxtaposed with semi colons. They describe the entity step by step and illustrate it, usually by linking users to an image (example 7) and by using metalanguage (*es decir*, *ejemplo*, etc.) that *always* explains the terms used in the definition. These are typically used in the general dictionaries of the portal and target native speakers and algorithms, i.e. these are used for creating patterns that will facilitate operations with *Write Assistant*, to which we will refer below:

balance (English: balance)
Definición
informe financiero que analiza el activo, o bienes o valores de los que dispone la empresa, y el pasivo, u obligaciones económicas de la misma como deudas, préstamos, etc., y su patrimonio neto, la diferencia entre el activo y el pasivo de la empresa; este análisis permite conocer la situación económica y financiera de una empresa o entidad en un momento determinado; se usa en economía
https://upload.wikimedia.org/wikipedia/commons/thumb/a/aa/Balances.JPG/395px-Balances.JPG

Example 7: Example of *special definitions* in the *Diccionarios Valladolid UVa* (Spanish general dictionaries).

4. *Descriptive definitions*: those which describe the meaning of qualitative lemmas, e.g. adjectives. There are two defining styles: a full Cobuild-style sentence or a chunk of words starting with "que" (English which or that) followed by one or more clauses. In both styles, we usually use explanations preceded by "i.e.", "esto es", or examples preceded by "e.x." with the aim of including denotative meaning (examples 8 and 9). The former type mainly targets human translators whereas the latter is addressed at native speakers and algorithms, as in the case of the previous definition style:

biannual
Definición
Biannual means half-yearly, i.e. every six months.

Example 8: Example of *Descriptive definitions* in the *Diccionarios Valladolid UVa* (English specialised dictionaries).

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

zazoso (English: stuttering)
Definición
que hace referencia a la persona que tartamudea, esto es, que tiene problemas para leer o hablar por repetir sílabas o pronunciar de forma entrecortada

Example 9: Example of *Descriptive definitions* in the *Diccionarios Valladolid UVa* (Spanish general dictionaries).

5. *Action definitions:* those which describe processes, e.g. verbs. These typically consist of one or more activities that interact to achieve a result, and they are also accompanied with exemplification (example 10); here both human speakers and algorithms are targeted:

abet
Definición
relacionarse con alguien con el único fin de lograr que dicha persona haga algo que no debería hacer, normalmente cometer un acto delictivo
equivalente
inducir

Example 10: Example of *Action definitions* in the *Diccionarios Valladolid UVa* (English–Spanish bilingual general dictionaries).

6. *Function definitions:* those which generally describe the meaning, if any, and function (usually) of lemmas such as adverbs, prepositions, interjections, and so on. We typically use chunks of words joined by semi-colons and target both human speakers and algorithms (example 11):

as at
Definición
expresión que indica un hecho que ocurre en un momento determinado; se utiliza normalmente en finanzas; se utiliza para conectar dos sintagmas, es decir un grupo de palabras que constituyen una unidad de funcionamiento, como por ejemplo "capaz de leer una novela", que es un sintagma adjetivo
equivalente
como en

Example 11: Example of *Function definitions* in the *Diccionarios Valladolid UVa* (English–Spanish bilingual general dictionaries).

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

7. *Pattern definitions:* those which describe the meanings in terms of semantic types (see Hanks' *Pattern Dictionary of English Verbs*). We use this defining style for explaining expressions, e.g. idioms, quotations, proverbs, and so on. We use chunks of words followed by a semantic type of the expression and its translation in bilingual dictionaries. The targets here are human speakers and algorithms (example 12):

   play one's ace card
   Definición
   emplear una persona su mejor recurso con intención de conseguir ventaja en una situación
   equivalente
   jugar su mejor baza
   - Someone plays his or her ace card
   - Alguien juega su major baza

Example 12: Example of *Pattern definitions* in the *Diccionarios Valladolid UVa* (English–Spanish bilingual general dictionaries).

8. *Equivalents:* All the bilingual dictionaries have one equivalent per definition, as shown in examples 10, 11 and 12.

In addition, our ad-hoc typology has allowed us to equip our dictionary portal with technologies for using two innovative types of search systems, both of which are currently working with the general Spanish dictionaries and the English–Spanish bilingual dictionaries. The first system is a WordFinder, i.e. a search system that explores the meaning and lexical remarks fields for concepts, i.e. search "definitions". Users can adapt their search with Boolean operators. For example, a Spanish user may look up American poisonous serpents by writing "serpiente – veneno + Hispanoamérica" in the search engine. This will retrieve dictionary articles for more than 100 words, e.g. the dictionary articles for *macaurel*, *boa común*, *mazacuata*, *tragavenado*, and *Boa constrictor imperator*. All these are non-poisonous serpents that live in Latin America. If the users changes the search string for "serpiente + veneno + Hispanoamérica", more than 100 dictionary articles will also be retrieved, but in this case poisonous serpents such as *serpiente de cascabel*, *crótalo*, *víbora de foseta*, and so on. As previously indicated, this system works by exploring the meaning and lexical remark fields, and can find the concepts because of the *special definition* type used (example 13):

macaurel (English: large poisonous snake)
Definición
serpiente de la familia de los boidos o boas que habita en Venezuela y otras zonas de América Central y del Sur; es una serpiente

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

> de hábitos nocturnos y no venenosa; es similar a la boa constrictor llamada tragavenado en Venezuela, pero más pequeña (de hecho, a menudo se considera que la tragavenados y la macaurel son el mismo tipo de serpiente); de color marrón, con manchas en forma de "H", musculosa y grande (puede llegar a medir cuatro metros); es temida por sus constantes ataques a las aves de corral y animales pequeños; habita, fundamentalmente, en las zonas cálidas y bajas de Venezuela

Example 13: Definition of *macaurel* in the *Diccionarios Valladolid-UVa* (Spanish general dictionary)

The second system also searches in the meaning field of the lexicographic database of the Dictionary Writing System, i.e. search "definitions". It is employed when a user logs on to *Write Assistant*, software developed to help Spanish native speakers to write English texts, and does not know an English word (Fuertes-Olivera and Tarp 2020). Examples 4, 6, 7, 9, 10, 11, 12 and 13 illustrate definitions that are very different from those existing in Spanish general dictionaries and English–Spanish bilingual ones. Our definitions will be particularly useful for the *Write Assistant*, which will be *mainly* fed by the *Diccionarios Valladolid-UVa*. *Write Assistant* is currently using statistics and will use machine learning and neural networks technologies for displaying patterns that will offer users of *Write Assistant* the most regular possibilities (Fuertes-Olivera and Tarp 2020). It basically works as follows: a Spanish native speaker is writing in English and he or she does not know or remember an English word. He or she can then write in Spanish and *Write Assistant* will offer them English possibilities, all of which are ordered according to frequency and possibility patterns that will be created with machine learning and neural network technologies that "learn" by analysing huge amounts of data; therefore, our definitions are long and complete. In this regard, although they are included in general dictionaries, the definitions in the examples above are basically similar to *terminological definitions*: they contain a lot of data, are very precise, and are linked to good contexts (chunks of words, collocations and examples) and images (e.g. facial recognition programs work with images). For instance, *coste atribuido* (example 4) is not present in *WordReference.com* (Spanish–English), *Diccionario de la Lengua Española* (RAE 2014), *Diccionario de Contabilidad Inglés–Español/ Español–Inglés* (Sánchez 2003), *Diccionario de uso del español* (Moliner 2007), nor *Diccionario del español actual (*Seco et al. 2011). Meanwhile, *zazoso* (example 9) is defined as *tartamudo* (*Diccionario de la Lengua Española)* and as *tartamudo* or *ceceoso* in the *Diccionario de uso del español* (Moliner 2007). The other dictionaries do not include this word. In the same vein, *macaurel* is poorly defined in the *Diccionario de la Lengua Española* (example 14):

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

macaurel (English: large poisonous snake)
Serpiente de Venezuela, no venenosa y parecida a la tragavenado, pero de menor tamaño.

Example 14: Definition of *macaurel* in the *Diccionario de la Lengua Española*

To sum up, the ad-hoc typology of definitions is associated with certain topics that are currently addressed in *terminology*, whilst offering a proposal for working with tools such as *Write Assistant*, and illustrating the economic value of lexicographic and terminological data in the data-driven economy associated with digital knowledge.

5. Conclusion

The *Diccionarios Valladolid-UVa* represent an attempt to make lexicographic and terminological activities profitable. This combines the tenets of the Function Theory of Lexicography (dictionaries are nothing more and nothing less than reference tools covering language, facts and things) with developments in business management. In this context, the focus is especially on the use of disruptive technologies for "attacking" the market dominance of incumbents and increasing the customer base of a product or service, together with the coming of age of the data-driven economy, which values the use of huge amounts of data for establishing patterns that can be used for many different activities, e.g. for upgrading existing language tools such as writing assistants. Both theoretical standpoints are based on two broad assumptions. Firstly, we have assumed that the concept of an integrated dictionary portal will allow us to include as many lemmas as possible, without limiting them and their lexicographic treatment to existing dictionary typologies, e.g. to differences between general and specialised dictionaries. Secondly, our dictionary entries are comprised on average of 200 words plus links to open data per sense. This makes them very different from existing dictionaries which as a rule use around 30 words and almost no link per sense (Fuertes-Olivera, Tarp and Sepstrup 2018). Of particular interest is the definition employed for describing each lemma; we have created an ad-hoc typology of definitions that is geared towards allowing AI technologies to develop patterns that will greatly improve the construction of assistants and similar language technologies. Already existing definitions found in monolingual and bilingual online dictionaries are of little use in this area.

Both developments explain the confluence of terminology and lexicography commented on at the beginning of this article. Following Bowker's Table 1, the following aspects may be appreciated regarding the *Diccionarios Valladolid UVa*: lexicographers, terminologists, translators and experts in the field work side by side; we analyse both words and terms in a similar fashion; we cover general and specialised domains; users can also search for domains (neither an alphabetical nor a thematic approach is necessary in online dictionaries); we

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

equip *all* our lemmas with up to 67 different lexicographic data; our intended users are professionals and organisations interested in and with the resources for using these dictionaries in combination with other tools, e.g. assistants.

We have illustrated the convergence of both disciplines with an ad-hoc typology of definitions. This typology is not based on any specific theory but on methodological issues (it facilitates daily work), and on the assumption that definitions such as those of examples 4, 6, 7, 9, 10, 11, 12 and 13 can be used with AI technologies. Results carried out in several countries using a Test Driven Development methodology are confirming our initial intuition that software such as *Write Assistant* can be of much use and that its users are not interested in differentiating between lexicography and terminology but in answering their queries in the fastest and easiest way possible.

References

Dictionaries and Tools

Diccionario de la Lengua Española. 2014. Real Academia Española. 2014. *Diccionario de la lengua española.* 23rd edition. Madrid: Espasa. http://www.rae.es/. Accessed 18 September, 2020.

Hanks, Patrick. 2000. *Patterns Dictionary of English Verbs.* http://pdev.org.uk/#about_cpa. Accessed 18 September, 2020.

Lexico.com. https://www.lexico.com/. Accessed 18 September,2020.

Moliner, Maria. 2007. *Diccionario de uso del español.* Third Edition. Madrid: Gredos.

OWID: http://www.owid.de/. Accessed 31 January, 2020.

Sánchez, Nora. 2003. *Diccionario de contabilidad. Inglés–español/español–inglés.* Hoboken: John Wiley.

Seco, Manuel, Olimpia Andrés and Gabino Ramos in collaboration with Mª Teresa de Unamuno, Juan Antonio Villafáñez and Carlos Domínguez. 2011. *Diccionario del español actual*. Madrid: Aguilar.

The Free Dictionary: http://www.thefreedictionary.com/. Accessed 18 September, 2020.

WordReference.com: http://www.wordreference.com/. Accessed 18 September, 2020.

Write Assistant: https://www.writeassistant.com/es. Accessed 18 September, 2020.

Other References

Boelhouwer, Bob, Hindrik Sijens and Anne Dykstra. 2018. Dictionary Portals. Pedro A. Fuertes-Olivera (Ed.). 2018: 754-766.

Bowker, L. 2018. Lexicography and Terminology. Pedro A. Fuertes-Olivera (Ed.). 2018: 138-151.

Christensen, Clayton. 2011. *Disruptive Innovation: The Christensen Collection.* Harvard: Harvard Business Review Press.

Christensen, Clayton, Michael E. Raynor and Rory McDonald. 2015. What is Disruptive Innovation? *Harvard Business Review,* December 2015. https://hbr.org/2015/12/what-is-disruptive-innovation. Accessed 18 September, 2020.

http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1598

Engelberg, Stefan and Carolin Müller-Spitzer. 2013. Dictionary Portals. Rufus, H. Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 1023-1035. Berlin/Boston: Mouton de Gruyter.

Fuertes-Olivera, Pedro A. 2015. Lexicographical Storing: A Key Lexicographical Task in the Era of the Internet. *Lexicographica* 31: 67-89.

Fuertes-Olivera, Pedro A. 2016. *European Lexicography in the Era of the Internet: Present Situations and Future Trends.* Plenary talk, Beijing, 2 December 2016. Talk sponsored by the Commercial Press and the Chinese Association of Lexicography.

Fuertes-Olivera, Pedro A. (Ed.). 2018. *The Routledge Handbook of Lexicography.* London: Routledge.

Fuertes-Olivera, Pedro A. 2019. Designing and Making Commercially Driven Integrated Dictionary Portals: The *Diccionarios Valladolid-UVa. Lexicography* 6(2): 21-41.

Fuertes-Olivera, Pedro A. and Ascensión Arribas-Baño. 2008. *Pedagogical Specialised Lexicography. The Representation of Meaning in English and Spanish Business Dictionaries.* Amsterdam/Philadelphia: John Benjamins.

Fuertes-Olivera, Pedro A. and Sven Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography.* Berlin/Boston: De Gruyter.

Fuertes-Olivera, Pedro A. and Sven Tarp. 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.

Fuertes-Olivera, Pedro A., Sven Tarp and Peter Sepstrup. 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the *Diccionarios Valladolid-UVa. Lexikos* 28: 152-176.

Harris, Roy and Christopher Hutton. 2007. *Definition in Theory and Practice. Language, Lexicography and the Law.* London/New York: Continuum.

Kilgarriff, Adam. 2006. Word from the Chair. Gilles Maurice de Schryver (Ed.). 2006. *DWS. Proceedings of the Fourth International Workshop on Dictionary Writing Systems, 5 September 2006, Turin, Italy (Pre-EURALEX 2006):* 7. Pretoria: $(SF)^2$.

Koeva, Svetla and Diana Blagoeva. 2013. The Dictionary Writing System Lexit and its Application in Bilingual Lexicography. *Cognitive Studies/Études Cognitives* 13: 57-65.

Müller-Spitzer, Carolin and Christine Möhrs. 2008. First Ideas of User-adapted Views of Lexicographic Data Exemplified on OWID and eLexiko. *Coling 2008. Proceedings of the Workshop on Cognitive Aspects of the Lexicon (CogALex 2008), Manchester, 24 August 2008:* 39-46.

Nielsen, Sandro. 2011. Function- and User-related Definitions in Online Dictionaries. Kartashkova, F.I. (Ed.). 2011. *Ivanovskaya leksikograficheskaya shkola: traditsii i innovatsii* [*Ivanova School of Lexicography: Traditions and Innovations*.] *A Festschrift in Honour of Professor Olga Karpova:* 197-219. Ivanovo: Ivanovo State University.

Rundell, Michael. 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.

Sager, Juan C. 1990. *A Practical Course in Terminology Processing.* Amsterdam/Philadelphia: John Benjamins.

Storjohann, Petra. 2018. German Lexicography in the Internet Era. Pedro A. Fuertes-Olivera (Ed.). 2018: 568-585.