Original Article

Designing a multilingual dictionary of genetic terms (English, French, German and Spanish) for the European Portal Eurogene and the International Scientific Community

CARMEN CUÉLLAR LÁZARO[1] , ANA M. MALLO LAPUERTA[2], ALFREDO CÓRDOVA MARTÍNEZ[3]

[1] *Faculty of Philosophy and Letters, University of Valladolid, Valladolid, Spain*

[2] *Faculty of Translation and Interpretation, University of Valladolid, Valladolid, Spain*

[3] *Faculty of Physiotherapy, University of Valladolid, Valladolid, Spain*

ABSTRACT

Cuéllar, C., Mallo, A.M. & Córdoba, A. (2014). Designing a multilingual dictionary of genetic terms (English, French, German and Spanish) for the European Portal Eurogene and the International Scientific Community *J. Hum. Sport Exerc., 9*(4), pp.773-781. In the framework of the DICGENETIC Project Applying Information Technologies to Designing an English-French-German-Spanish Dictionary of Genetic Terms, financed by the Regional Government of Castilla y León, the distinguished research group team Intersemiotics, Translation and New Technologies was assigned to design an online multimedia terminology dictionary to be integrated in the learning platform Moodle. The objective of this dictionary is to provide service to users of the European genetics portal Eurogene and also to the international scientific community, with a subsequent electronic version and via the Internet. **Key words**: MULTILINGUAL TERMINOLOGY DICTIONARY; GENETICS; ICT; SPECIALISED LANGUAGE ENGLISH, FRENCH, GERMAN, SPANISH; EUROGENE.

Corresponding author. Faculty of Philosophy and Letters, University of Valladolid, 47011 Valladolid, Spain.
E-mail: carmen.cuellar@lesp.uva.es
Submitted for publication December 2013
Accepted for publication June 2014
JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202

doi:10.14198/jhse.2014.94.04

INTRODUCTION

Human genetics has become the fastest growing biomedical sciences area in the decades since 1990, when the most ambitious scientific project in history started: the Genome Project. Not only has it brought a wealth of knowledge but it has also produced multiple scientific concepts. Consequently, it has generated similarly numerous terms with which to define them.

To allow students and professionals to be suitably trained and to update human genetics content, new tools have been designed that make use of the new technologies. Among the most interesting and innovative tools is Eurogene (www.eurogene.eu). It was financed by the e-Content Plus Program, in whose Promoting Consortium the University of Valladolid has been involved as a Content Partner from day one and throughout its entire three-year lifespan. The Eurogene portal, soon to be known as Edugene, is a multilingual reference portal for learning about human genetics that can be used in different educational levels, from secondary education to genetics experts. More than 300,000 free open access teaching units are available in this web portal for anyone interested.

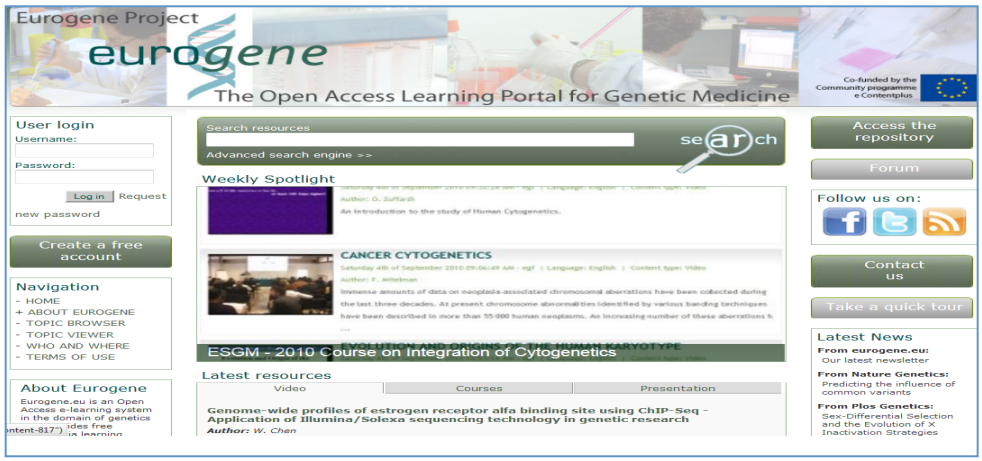

Figure 1. *Eurogene Homepage screenshot (www.eurogene.eu)*

One of the main difficulties encountered by users, both health professionals and students, when trying to get the best from a resource such as Edugene lies precisely in understanding some of the concepts defined with neologisms added to this field of knowledge. In this respect, terminology plays a central role in transferring specialised knowledge and is a crucial element in specialised texts (Arntz & Picht 1995, Alcaraz Varó 2000, Cabré 2004, Gómez & Vargas 2004).

The DICGENETIC Project was born in this context, as part of the University of Valladolid and financed by the Department of Education of the Regional Government of Castilla y León.

THE DICGENETIC PROJECT: AIMS

The primary aim of this interdisciplinary project is offering a dictionary of genetic terms in four of the main world languages, linked to the Edugene portal. It also focuses on conducting research on the contribution of information and communication technologies (ICTs) to get the best from the educational resources.

The project has the following specific aims:

• Increasing awareness about the importance of ICTs in genetic research.
• Involving the Castilla y León Region in promoting, for the international scientific community, the development of a tool that enables documentary search on genetics by applying multilingual lexical resources.
• Establishing unambiguous specific terms in every working language.
• Optimising the presence of the Region in the VII EU Framework Program through the e-Content Plus Program, a well-established international project.
• Filling the existing gap in the specialised field of genetic terminology, necessary for the scientific and social community.
• Strengthening collaboration among universities and other research centres in developing R+D+I policies.

These specific aims are established to promote ICTs in genetic research, to clarify the terminology in this field, to make it easier to acquire knowledge through this terminology tool, to derive definite meanings and to promote translation in a globalised society.

THE DICGENETIC PROJECT: HUMAN TEAM

The ten members of the research group Intersemiótica, Traducción y Nuevas Tecnologías (Intersemiotics, Translation and New Technologies) were the founders of the team. The group was rated as "Excellent" for its research, given its multidisciplinary approach, its efforts to increase collaboration among different fields of knowledge and its European dimension. Many other collaborators from different national and international bodies and institutions have joined the group since then; the Project currently has 47 researchers who cover the four working languages: English, French, German and Spanish.

The interdisciplinary nature of the Project is reflected in the team involved. The terminological and translation work that linguistic researchers and terminologists carry out has to be complemented by the documentary support and advice of experts in genetics (specifically human genetics, a subfield in which the dictionary falls). In other words, cooperation with the health sector is exceptionally important for our project and physicians and biologists are involved in its development.

Telecommunication engineers work to research and develop the prototype that will host the dictionary. They are also the group members in charge of building the database and subsequently publishing the online dictionary.

It is precisely the Eurogene representative in Spain and member of the DICGENETIC Project, Dr Juan José Tellería, who is in charge of conducting the research. Dr Tellería is a researcher in the Biology and Molecular Genetics Institute (IBGM is the Spanish acronym) at the University of Valladolid (http://www.ibgm.med.uva.es/).

The Translation and Interpretation School, the Faculty of Arts and the Advanced Technical School of Telecommunication Engineers of the University of Valladolid are joined by collaborators from other national institutions such as the University of Alicante and researchers from international institutions such as the Université d'Artois (France), the Universität Saarbrücken (Germany), the Friedrich-Schiller Universität Jena (Germany), the Rheinische Friedrich-Wilhelms-Universität in Bonn (Germany) and the Universitatea de

Vest (Romania). Further support is also provided from specialised centres, such as the Medical School of Soria and the aforementioned IBGM.

DICGENETIC PROJECT: METHODOLOGY

Terminology work is carried out throughout different phases (Vargas Sierra, 2008, idem, 2009). Specifically, the following sequence has been established for the DICGENETIC Project:

- Documentation (illustrative and terminological).
- Conceptual structuring of the working field from a terminology perspective.
- Working text selection (corpus building): Terminology extraction.
- Terms description: File building.
- Creating a terminology database (digitalising collected data).

In selecting working texts, the focus has been on choosing current specialised original sources (i.e., texts must be written in their authors' native language). To this extent, the texts we have gathered come from scientific papers, guidelines, etc. (Solari & Roubicek, 2007; Snustad & Simmons, 2010; Teufel, 2011; Serre, 2012; Schaaf & Zschocke, 2013).

Conceptual structuring of the working field takes place in a phase prior to text selection. Dividing into subfields has facilitated distributing the workload among various researchers. In doing so, each individual or group can specialise in a particular subfield and select and describe the terms of that particular subject. In line with this, we began our work in the field of eukaryotic cells. This was certainly a broad subject, so it was therefore broken into four subfields related to the composition of eukaryotic cells. The subfields are as follows: organelles, genes, proteins and nucleic acids.

Four working groups, made up of researchers covering the four languages, were established on the basis of these subfields. Their task was to look for specific documents in each language, to select working texts and to carry out terminology extraction.

- Group 1: Organelles
- Group 2: Genes
- Group 3: Proteins
- Group 4: Nucleic acids

The lists each group prepared were reviewed by Project member specialists and we are currently handling a total of 183 terms.

The terminology file is created according to the methodology set forward in ISO 10241:1992 and ISO 12620:1999 Standards. Both these standards are also reflected in the model card of TermEsp (Spanish Superior Council for Scientific Research) and in different types of data subject to being included in a terminology file (gathered in Cabré 1993, 279). These materials were the basis for determining which fields had to contain mandatory information and which were optional.

The card has to follow a structured pattern that makes it possible to delimit, synthesise and systematize a notion and then relate it to one or more denominations. Consequently, it should be divided in a number of predefined fields and data, which are geared to the purpose of every task. Such fields configure the

dictionary macrostructure. In the case of this Project and taking into account the fulfilment of its objectives, the information contained in the cards falls under the following three data categories:

• Management Data
• Linguistic Data
• Conceptual Data

Taking into account the linguistic and conceptual information is repeated in each working language, the card is then configured as follows:

Management Data

• Name / Project Number
• Language Code / Country Code
• Producing and Editing Data
• Institution
• Correlative Number
• Author

Linguistic Data

• Entry or Term: Denominating a concept deemed as highly reliable and more widely used.
• Term Source or Reference.
• Grammatical category: As the number of categories is small and to be repeated over and over, efforts were made to keep the data that could appear in this field consistent. To this end, an abbreviation list (m. for masculine noun; f. for feminine noun; n for neutral noun; v. or vtr. for transitive verb; vintr. for intransitive verb; vpr. for pronominal verb; adj. for adjective; adv. for adverb) was used.
• Context: Segments illustrating term usage. They provide information that can be very useful for delimiting an entry conceptually. There are different types of contexts:

- Testimonial Context: It indicates only that the term is in use but does not give further information about it.
- Defining Context: It provides accurate, useful and necessary information about the concept but it is not exactly a definition.
- Explanatory Context: It provides information about some of the core features of the concept contained in the definition.
- Associative Context: It allows associating the term with other concepts in the study field.
- Metalinguistic Context: It provides some etymological information or it indicates the gender, number and grammatical category of the term. It also provides a linguistic explanation of the word and its usage.

• Context source or reference
• Phraseology: Collocations (stable combinations of words) whose usage is preferred over other workable structures.
• Referrals: Neighbouring concepts. These are fields that are reciprocally related, whether by semantic equivalents, synonymy (terms and their geographic variants or dialects, symbols, initials, abbreviations,

scientific names or abbreviations) or by inclusion (hyperonyms, hyponyms or co-hyponyms) or by contrast (antonyms).
• Weighting mark or reliability code: It provides information about the standardisation of the term:

- Normative or standardised term
- Neologism pending approval
- Term pending review
- Rejectable term

Conceptual Data

• Subarea or Subfield: This section describes the conceptual structuring of the field, so a terminological tree has to be created in advance.
• Definition: The terminological definition describes the notion only in reference to the ongoing specialty domain. Likewise, it has to include the characteristics needed to distinguish the concept from other similar ones (Faber 2002). Some consistency criteria have been established in order to incorporate the definition.

- Lower case has to be used except for the initial letter.
- Only one clause applies (without internal stops).
- Both the first describer and the described term have to have the same grammatical category.
- It can never be a negative definition or defined by synonyms.
- Metalinguistic structures, such as "term used for…", "a noun that…", etc., should be avoided.

• Definition Source or Reference.
• Illustration: An image, a drawing, a video or a sound recording that refers specifically to the entry term. Usually a single illustration is used for all four languages but an exception is made with schemes establishing parts of an object, element or process.
• Illustration Source or Reference.
• Notes: Anything that seems interesting about the term, whether from the conceptual point of view or from the linguistic point of view.

In terminography, the cards are known as "extraction files". Any information found about every term is introduced in the cards as it is found. In the end, we select the most relevant data from what was collected and then enter those into the digital file.

The Project counts with the support of a collaborative Moodle platform that hosts the digital card. The platform is a highly useful tool for communication and interaction among the stakeholders involved. The platform is inserted in the Distance Learning Service of the aforementioned Intersemiotics, Translation and New Technologies Group and of the Infrastructures, Technologies, Telecommunication Service and Application Group (ITAST is the Spanish acronym). The Distance Learning Service has extensive experience in research on and the practical application of e-learning methodologies, systems and services

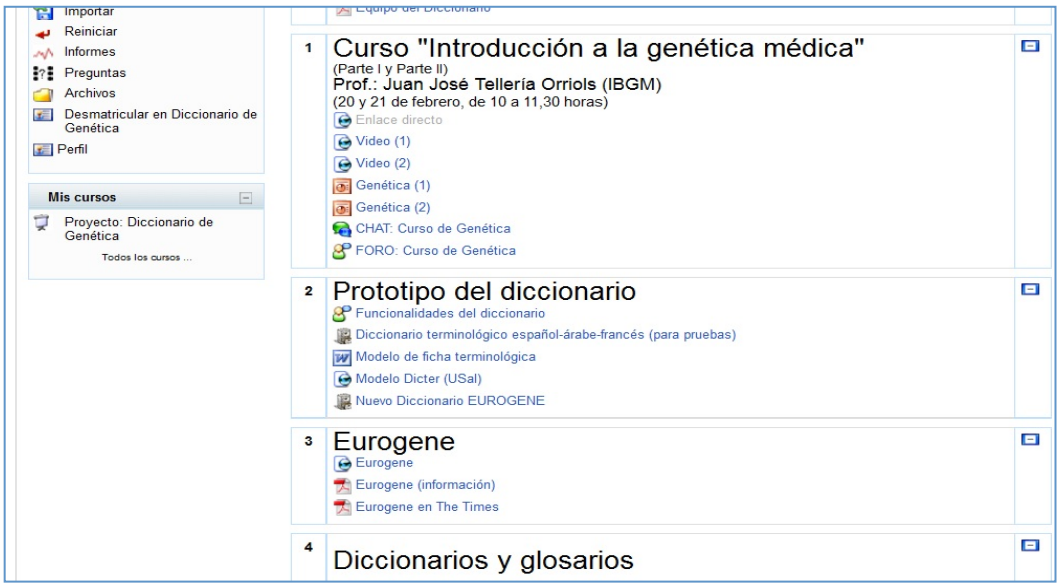

Figure 2. *DICGENETIC Project Moodle Platform Screenshot*

The following is a screenshot from the "chromosome" terminological card located in this platform:

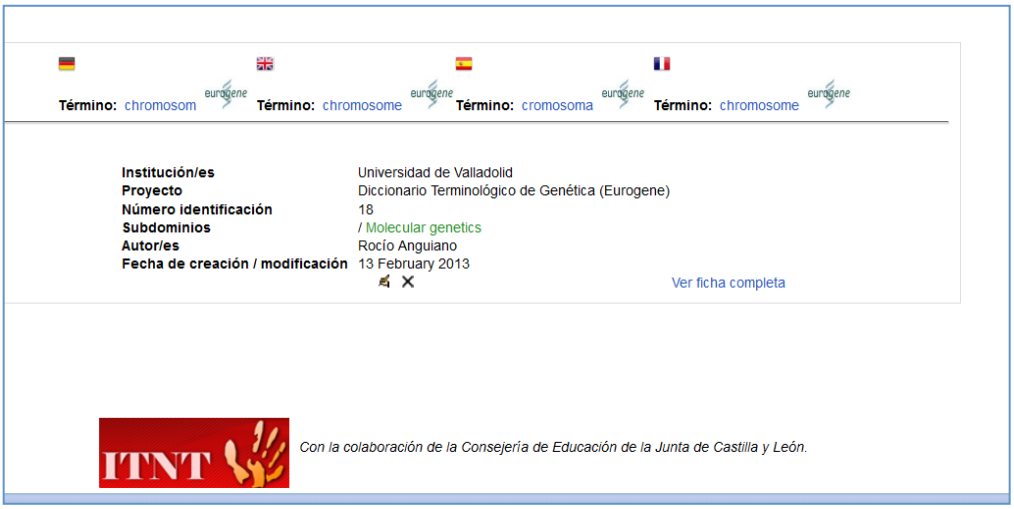

Figure 3. *Screenshot of the "chromosome" term card*

In the course of the DICGENETIC Project, the research team received two training courses. On the one hand, there was the course on "Terminography Applied to Designing a Genetic Terminological Dictionary", which was led by a member of the terminology expert group. Its purpose was to set the theoretical and practical foundations to unify the selection, planning and description criteria for the terms comprising the multilingual terminological dictionary of Genetics. The second course was "Introduction to Medical Genetics", which was given by a member of the Genetics expert group. It was aimed at teaching the basic concepts to facilitate designing the card. The Moodle platform has also provided access to very interesting material for Project development (links for dictionaries, glossaries, etc.).

The tremendous work involved in the coordination of an interdisciplinary Project of this nature was carried out by the main Project researcher, Dr Antonio Bueno García.

DICGENETIC PROJECT: CONCLUSIONS. THE DICTIONARY FUNCTIONS

The online multimedia terminology dictionary will be integrated in the Moodle learning platform in order to provide service to users of the European genetics portal Eurogene and to the international scientific community, with a subsequent electronic version and via internet. In other words, the new tool will be integrated, not only in Moodle but also in the Eurogene portal, to allow users a two-way flow between Moodle courses and the portal. In doing so, access to a dictionary created in a public Moodle course will be possible from Eurogene. Likewise, it will also be possible to access Eurogene resources from Moodle courses that include the module.

In order to facilitate access to and the development of the dictionary to be created in the Project framework, the dictionary should be digitalised so that the concepts chosen and all the information linked to them in the four working languages will be stored in a database accessible from a web interface. It will work in two different ways. First, as an administrator, to allow the dictionary authors to incorporate different defined terms and manage all the information linked to them, which can be either modified or removed if necessary. Second, the interface can also function in user status, allowing science professionals working in the project framework to search for information contained in the dictionary using different criteria.

The final product will consequently be a multilingual digital dictionary: it will be possible to link information in different languages and formats (including text, image, audio and video) to every term. Lastly, it is also expected to include training packages, along with the terminological card. Actually, the instructional capacity of the dictionary will be an added value of the research.

We conclude this article, hoping the dictionary will soon see the light of day and that, in this increasingly globalised world, the dictionary will serve as a tool to enhance communication internationally.

REFERENCES

1. Alcaraz, E. (2000). *El inglés profesional y académico*. Madrid: Alianza Editorial.
2. Arntz, R., & Picht, H. (1995). *Introducción a la terminología*. Madrid: Fundación Germán Sánchez Ruipérez, Ediciones Pirámide.
3. CABRÉ, M.T. (1993). *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártida
4. CABRÉ, M.T. (2004). "La terminología en la traducción especializada". In C. GONZALO GARCÍA & V. GARCÍA YEBRA (EDS.), *Manual de documentación y terminología para la traducción especializada* (pp. 89-122). Madrid: Arco/Libros, 89-122.
5. Faber, P. (2002). "Terminographic definition and concept representation". In B. MAIA, J. Haller, & M. Ulyrich (Eds.). *Training the Language Services Provider for the New Millennium.* Porto: Universidade do Porto, 343 354.
6. Gómez, A., & Vargas, C. (2004). "Aspectos metodológicos para la elaboración de diccionarios especializados bilingües destinados al traductor". In L. GONZÁLEZ & P. HERNÚÑEZ *El español, lengua de traducción. II Congreso internacional.* Brussels: ESLEtRA, 365-398.
7. Schaaf, C.P. & Zschocke, J. (2013). *Basiswissen Humangenetik*. Springer.
8. Serre, J.L. (2012). *Génétique. Théorie, analyse et ingéniérie*. Paris: Dunod.
9. Snustad, D.P. & SIMMONS, M.J. *Principles of Genetics*, sixth edition, New York: John Wiley, 2010.

10. Solari, A.J., & Roubicek, M. (2007). *Genética humana: Fundamentos y aplicaciones en medicina*. Buenos Aires: Médica Panamericana.
11. Teufel, A. (2011). *BASICS Humangenetik*. Elsevier.
12. Vargas, C. (2008). "La sistematización terminográfica: una propuesta metodológica para la elaboración de diccionarios traductológicos". In *Actas del X Simposio Iberoamericano de Terminología*, Montevideo, Uruguay (CD-ROM).
13. Vargas, C. (2009). "Do-it-yourself IT for Terminology o experiencias de bricolaje informático en la elaboración de diccionarios terminológico". In *Puntoycoma*, 115s, Brussels; Luxemburg: Directorate-General for Translation of the European Commission, 42-49.