European Journal of Clinical Nutrition (2010) 64, S44–S48
 0954-3007/10
www.nature.com/ejcn

ORIGINAL ARTICLE

LanguaL Food Description: a Learning Process

JD Ireland and A Møller

Danish Food Information, Roskilde, Denmark

The European Food Information Resource (EuroFIR) Network of Excellence (http://www.eurofir.eu) aims to provide validated food composition data (FCD) from European food composition databanks. However, the network covers 27 countries with different languages and food cultures, making comparisons difficult. Moreover, in 2005, only 25% of the databases included international food classification or food description. To overcome some of the challenges, it was decided to use the LanguaL thesaurus (http://www.langual.org) to index (systematically describe) and link the foods in European food composition databases (FCDBs). To facilitate food indexing, LanguaL Food Product Indexer software was introduced and several short (1–2 days) food indexing courses for FCD compilers from all participating countries were organised.
Feedback between the LanguaL Technical Committee and the FCD compilers allowed the latter to improve their food indexing skills. In turn, the compilers proposed new descriptors and translations for the thesaurus. The result was a set of more than 26 000 foods in national databases and 2360 foods in specialised data sets that were LanguaL indexed and thus able to be linked to the EuroFIR network.
Both the EuroFIR network and the individual FCD compilers benefit from standardised food description, allowing foods to be linked and compared across borders and language barriers. The LanguaL thesaurus has, in turn, benefited from the expertise of the FCDB compilers from different cultures.
European Journal of Clinical Nutrition (2010) **64,** S44–S48; doi:10.1038/ejcn.2010.209

Keywords: LanguaL; thesaurus; food description; EuroFIR; food composition database

Introduction

The European Food Information Resource (EuroFIR) Network of Excellence (http://www.eurofir.eu) aims to build and disseminate a comprehensive, coherent and validated databank system providing a single, authoritative source of food composition data (FCD) in Europe for nutrients and newly emerging bioactive compounds with positive health effects. The EuroFIR eSearch facility is a virtual platform of interconnected, online national and specialised databases/data sets. A distributed configuration, using client-mediator server applications, interlinks the data sets and external systems relevant for the food composition area (Møller *et al.*, 2007). The EuroFIR network covers 27 countries with different languages and food cultures, which constitutes a special challenge for food composition data interchange and comparison.

Work on food composition databases (FCDBs) has traditionally been centred on the publication of printed national or specialised food composition tables, with limited space for data documentation. The level of available detail in these tables is thus generally not sufficiently specific to be used as input for compilers in other countries. Moreover, as databases use different methods of classifying and describing foods, it is difficult to exchange data between countries or even between organisations within the same country.

When the EuroFIR project started in 2005, it carried out a survey of food classification and description in European FCDBs. The results of the survey were not surprising and showed that the situation in Europe was similar to that of the biblical Tower of Babel: foods in databases were invariably described by *ad hoc* classifications and by food names in national languages; English translations and scientific names were not systematic; only 25% of FCDBs included international food classification or food description (Ireland, 2005).

Materials/subjects and methods

One of the major uses of FCDBs is the assessment of dietary intakes. Foods consumed by individuals are determined using dietary recalls or records and are then 'matched' to

Correspondence: Dr JD Ireland, Danish Food Information, Borgediget 12, Roskilde 4000, Denmark.
E-mail: ji@danfood.info

foods in a database so that the daily intakes of nutrients may be estimated and assessed. The accuracy of the 'matching' process depends on how well the foods are named and described by the subject and how the foods are described in the FCDB. The more clearly and completely that foods are described, the better the matches and the more accurate the assessment of dietary intake will be (Pennington and Butrum, 1991).

The preparation of reliable data on food requires precise nomenclature and detailed description of foods. Even data of good quality can be a source of error if they are derived from foods that are not clearly defined. Furthermore, it is difficult to exchange data on foods, or to understand and compare dietary patterns and nutritional status for different countries or individuals, without coherent description and classification of foods in databases (Ireland and Møller, 2000). To allow data interchange, food description should make use of standardised vocabularies (thesauri), that is, sets of agreed or standardised terms, maintained and published by an authoritative body. Examples of standardised vocabularies are ISO names of countries and languages, Codex Alimentarius food classifications and the LanguaL thesaurus.

However, food classification is not sufficient. Food classification systems have been created for different purposes and reflect different legislations and needs (Ireland and Møller, 2000). Their purpose is to aggregate foods with similar characteristics. For example, the Codex General Standard for Food Additives classification (Codex Alimentarius, 2009) lists food groups according to the additives they may contain, whereas the Codex Committee on Pesticide Residues classification (Codex Alimentarius, 2006) is residue and contaminant driven. A food description system, on the other hand, is multidimensional and allows information to be recorded on other significant characteristics of a food. As the EuroFIR network needs to document data in a multilingual, multicultural environment, it chose the LanguaL thesaurus to index (systematically describe) and link European databases at the food level (Becker *et al*., 2007).

LanguaL ('*Langua aLimentaria*' or 'language of food') is a multilingual, facetted thesaurus created to describe foods in a systematic manner. The thesaurus was initiated in the late 1970s (McCann *et al*., 1988) by the Center for Food Safety and Applied Nutrition of the United States Food and Drug Administration and then further developed in collaboration with the US National Cancer Institute and European partners. Since 1996, the European LanguaL Technical Committee has administered the thesaurus and made it available online (http://www.langual.org). Using the LanguaL thesaurus, each food in a database can be described by a set of standard, controlled terms arranged in 14 facets characteristic of the nutritional and/or hygienic quality of a food, such as the biological origin, technological treatments, packaging, as well as methods of cooking and preservation (Møller *et al*., 2008).

The first challenge in using the LanguaL thesaurus in the EuroFIR network was the inclusion of a common food classification system that could be used by all European FCDB compilers. When the EuroFIR project began, the food classification systems present in the LanguaL thesaurus were based on US legislation (for example, Code of Federal Regulations) or specific regulation regarding food safety (for example, Codex classification of food and feed commodities, Confederation of the Food and Drink Industries (CIAA) food classification for food additives). EuroFIR therefore created and included a food classification system for accessing European FCDBs, by combining the most common food groups or classes of these databases (Ireland and Møller, 2006). These FCDB food groups were also mapped to food classification systems created for European food consumption or food availability surveys: Eurocode2 (http://www.eurofir.org/eurocode/), DAFNE (http://www.nut.uoa.gr/english/dafne/DafneEN.htm), EPIC (http://epic.iarc.fr/) (Slimani *et al*., 2000) and EFG (Ireland *et al*., 2002). The objective was to create a food classification system for FCDBs that was compatible with COST Action 99 and EFCOSUM recommendations (Ireland *et al*., 2002) for food consumption surveys, without constituting a recommendation for the latter. The resulting EuroFIR food classification system for accessing FCDBs was then included in the LanguaL thesaurus (Møller and Ireland, 2007).

Another challenge in using LanguaL to index foods in all European FCDBs stems from the complexity of the thesaurus (over 5000 descriptors in 14 facets). It was thus decided to provide FCDB compilers with a tool to facilitate food indexing. Specific software was developed for this: the LanguaL Food Product Indexer (FPI). The FPI software incorporates hierarchy (tree structure), definitions, synonyms and scope notes of descriptors in the LanguaL thesaurus, plus a number of already indexed data sets as examples of food indexing (Figure 1). The LanguaL FPI software and accompanying documents (FPI tutorial, LanguaL reports and indexed food files) have been made available for download from the LanguaL website (http://www.langual.org/langual_downloads.asp).

Between 2005 and 2009, several short (1–2 days) food indexing courses were organised for food composition data compilers (Table 1). The courses covered food nomenclature and authoritative sources of scientific names; international systems of food classification and description; history and principle of the LanguaL thesaurus; structure of the LanguaL thesaurus, with definitions and indexing rules for the different facets characterising foods. Several examples of food indexing were also carried out with the participants. Course participants received compact discs or USB flash memory disks with LanguaL documents, the FPI software and a selection of indexed food lists to help them index their national foods.

After the first food indexing course (November 2005), the European FCDB compilers were instructed to index about 500 priority foods from their databases by March 2006 (first round of food indexing) and 1000 foods by the end of 2006 (second round).

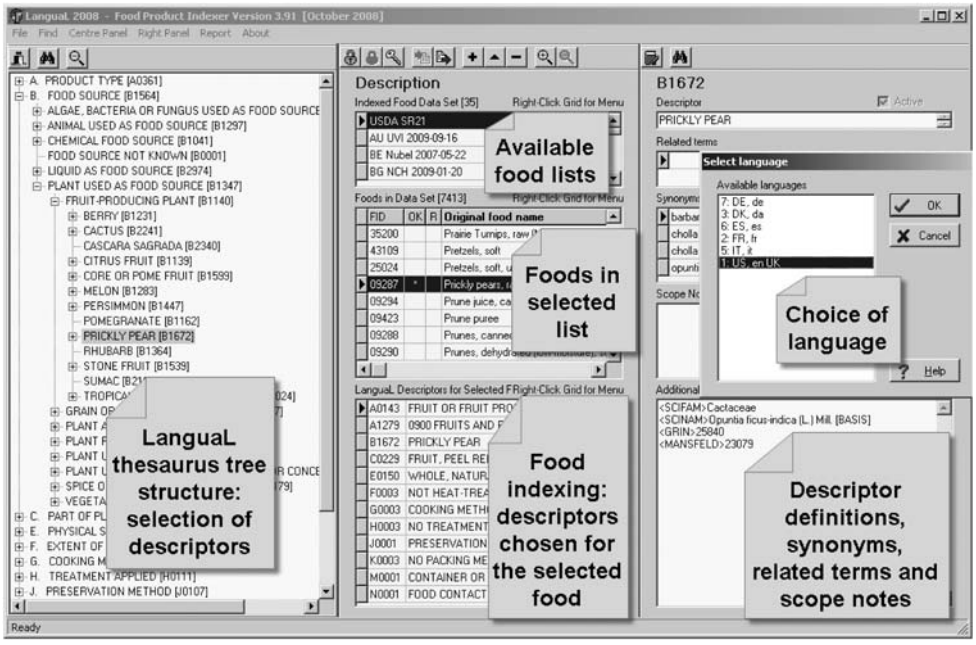

Figure 1 LanguaL food product indexer software.

Table 1 Food indexing training courses

Year	*Course*	*No. of days*	*No. of participants*
2005	Food indexing course, Wageningen (NL)	2	24
2006	FoodComp, Bratislava (SK)	1	30
2007	Food indexing course, Paris (FR)	2	29
2007	FoodComp, Wageningen (NL)	1	29
2008	FoodComp, Bratislava (SK)	1	26
2008	USDA/FDA workshop, Washington DC (USA)	1	30
2009	Latinfoods workshop, Tucumán (AR)	1	24
2009	BaSeFood workshop, Belgrade (SR)	1	28

Results

The characteristics of foods described by LanguaL can have a major impact on the assessment of the levels of nutrients and other dietary components in foods. Food composition can vary widely according to technological treatments (adding, substituting or removing components or modifying the food, for example, through fermentation), cooking methods (for example, frying or boiling), preservation methods (for example, sterilisation, drying) and other treatments. Food description accompanying data should therefore be as accurate and complete as possible. The quality of systematic food description can be assessed according to the following criteria (Ireland and Møller, 2007):

- Reproducibility: the extent to which different people index identical foods in the same way. Reproducibility can be evaluated by comparing the results of food indexing by different compilers.
- Correctness: the extent to which different people index foods correctly (two people may index the same foods identically but both may be wrong). Correctness can be evaluated by comparing the results of food indexing by the different compilers with those of experts in LanguaL.
- Completeness: the extent to which all the characteristics of interest to a food scientist or a nutritionist have been indexed. Completeness of LanguaL indexing is to a large extent assured by FPI software, which verifies that all facets have been indexed for each food in a list.

The compilers' food indexing was evaluated after the first two rounds (2006 and 2007). After the first round, about 1000 foods out of 20 000 were found to present indexing problems (for example, determining food source, definition of physical state at 20 °C). In the second round of food indexing, $<2\%$ of foods required corrections, there were few serious indexing problems and most compilers worked in a similar way. Nevertheless, the completeness and accuracy of indexing largely depends on the compiler's knowledge of food processing and/or food systems. Compilers work within their own national FCDBs and may not have access to technical experts who may be familiar with food processing operations. By allowing compilers to copy preexisting descriptions to their foods, the FPI software provides them with some of this technical information and helps to harmonise food description. These results show that the two food indexing courses were successful and that FPI software is easy to use (Ireland and Møller, 2007).

Feedback between the LanguaL Technical Committee and FCDB compilers also allowed the latter to improve their indexing skills. Quality assessment of the compilers' food indexing was necessarily based on the English translations of food names, and some of the irregularities reported back to the compilers may have come from a misunderstanding of the foods described. It is very often difficult to translate food names of specific national origin into English, which is a foreign language for all partner countries, except for Great Britain. Some English translations of food names in national databases were also improved when the LanguaL description was found to be more accurate. For example, a food item translated as 'Bream, raw' and indexed as *Cusk* [B2143] actually turned out to be cusk (*Brosme brosme* Ascanius, 1772); thus, the common name could be corrected in the national database. In the second round, some FCDB compilers included more information in their English food names (for example, 'Makrud—semolina cake with dates'), distinguished between homemade and commercial products and indicated brand names where appropriate. This is a good practice and should be generalised in FCDBs, as the food name can contain additional information not covered by the standardised indexing.

The feedback from FCDB compilers also revealed that some improvements were needed in the LanguaL thesaurus, such as new descriptors for a range of European plant and fish species, updated Scope Notes, synonyms and additional information for some other descriptors (Møller and Ireland, 2007). Moreover, compilers provided translations of the thesaurus in several European languages, in addition to English, namely, Danish, French, German, Italian and Spanish (Møller and Ireland, 2008).

The final result of the LanguaL indexing project was a set of more than 26 000 foods in national databases (Table 2) and 2360 foods in specialised data sets (Table 3) that were LanguaL indexed and thus able to be searched in the EuroFIR network's data facilities (Møller *et al.*, 2007).

Table 2 National data sets indexed with LanguaL

Indexed data sets	*No. of foods*	*Indexed data sets*	*No. of foods*
Czech Republic UZEI	77	Turkey TUBITAK	974
Lithuania NNC	135	Germany BfEL-BLS	1034
Denmark DFI_NDS	339	Ireland UCC	1050
Switzerland SwissFIR	452	Italy CSPO	1052
Austria UVI	514	Serbia IMR	1141
Sweden NFA	467	Norway MVT	1188
Latvia FVS	523	France AFSSA	1346
Spain UGR	658	Slovakia FRI	1400
Bulgaria NCH	828	Denmark DTU_FDB	1546
Greece NUKA	901	Great Britain IFR	1703
Italy INRAN	909	Israel BGU	1925
Poland NFNI	932	Hungary UB	2078
Belgium Nubel	944	Finland FINELI	2093
Iceland ISGEM	946	Netherlands RIVM-NEVO	2309
Portugal INSA	962	United States USDA (SR21)	3489

Discussion

Both the EuroFIR network and individual compilers benefit from standardised food description, allowing foods to be linked and compared across borders and language barriers. The results of the food indexing evaluation provided examples of how real problems in food naming and food indexing can be solved. The evaluation, moreover, made food naming and food indexing more precise and consistent among partners in the EuroFIR network. As a result, record retrieval based on food description can be more reliable. Moreover, standardised food description leads to better comparability and communication among technical experts for a specific use of the data, for example, for food consumption surveys.

An important outcome is the possibility of using LanguaL food indexing to match foods cited in national consumption surveys with records in FCDBs. The more clearly and completely that foods are described, the better the matches and the more accurate the assessment of dietary intake will be. An information system based on LanguaL descriptors could thus be incorporated in software to assist food coding within a dietary survey (Turrini *et al.*, 2009).

Furthermore, using LanguaL indexing for foods in FCDBs also enhances data retrieval results in data information systems such as the EuroFIR eSearch platform, because the LanguaL indexing allows for more precise descriptions of foods than the often ambiguous food names themselves.

The LanguaL thesaurus has, in turn, benefited from the expertise of the compilers. The latter have proposed new descriptors based on European foods and have provided translations into national languages, which have since been incorporated into the LanguaL FPI software. The

Table 3 Specialised data sets indexed with LanguaL

Data set	*Website*	*No. of foods*
EuroFIR BASIS: bioactive substance in foods	http://ebasis.eurofir.org/	300
Phenol-explorer database	http://www.phenol-explorer.eu/	452
GEMS/Foods: WHO/FAO Global Environmental Monitoring System, using Codex CCPR classification	http://www.who.int/foodsafety/chem/gems/en/index.html	1549
INFORMALL: information on allergenic foods	http://foodallergens.ifr.ac.uk/	61

food indexing evaluation has also promoted discussion among compilers and data users to determine correct food description.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This work was completed on behalf of the EuroFIR consortium and funded under the EU 6th Framework Food Quality and Safety Programme (FOOD-CT-2005-513944). We are grateful for the support of all EuroFIR partners.

References

Becker W, Unwin I, Ireland J, Møller A (2007). *Proposal for structure and detail of a EuroFIR standard on food composition data I: Description of the standard*. EuroFIR report.

Codex Alimentarius (2006). Codex Classification of Foods and Animal Feeds. Available at: ftp://ftp.fao.org/codex/ccpr38/pr38CxCl.pdf (accessed March 2010).

Codex Alimentarius (2009). Codex General Standard for Food Additives. Available at: http://www.codexalimentarius.net/gsfaonline/foods/index.html(accessed March 2010).

Ireland J (2005). Food Classification and Description in European Food Composition Databases. EuroFIR report D1.6.2.

Ireland J, Møller A (2000). Review of international food classification and description. *J Food Compos Anal* **13**, 529–538.

Ireland J, Møller A (2006). EuroFIR workshop and draft recommendations for standard food classification and description systems for use in European food composition databases. EuroFIR report D1.6.4.

Ireland J, Møller A (2007). Results of Food Indexing by EuroFIR Compilers Network. *EuroFIR report* D1.8.5.; Report on food record retrieval using proposed description and classification. *EuroFIR report D1.8.1*.

Ireland J, van Erp-Baart AMJ, Charrondiere UR, Møller A, Smithers G, Trichopoulou A (2002). Selection of a food classification system and a food composition database for future food consumption surveys. *Eur J Clin Nutr* **56**(Suppl 2), S33–S45.

McCann A, Pennington JA, Smith EC, Holden JM, Soergel D, Wiley RC (1988). FDA's factored food vocabulary for food product description. *J Am Dietetic Assoc* **3**, 336–341.

Møller A, Ireland J (2007). LanguaL 2007—Documentation of changes from version 2000. *EuroFIR Technical Report* Danish Food Information.

Møller A, Ireland J (2008). LanguaL 2008—Multilingual Thesaurus (English–Danish–French–German–Italian–Spanish). EuroFIR Technical Report D1.8.21d. Danish Food Information.

Møller A, Ireland J, Smith E (2008). LanguaL 2008—Introduction. EuroFIR Technical Report D1.8.21a. Danish Food Information.

Møller A, Unwin ID, Becker W, Ireland J (2007). EuroFIR's food databank systems for nutrients and bioactives. *Trends Food Sci Tech* **18**, 428–433.

Pennington JAT, Butrum RR (1991). Food description—LanguaL and Taxonomy. *Trends Food Sci Tech* **2**, 285–288.

Slimani N, Ferrari P, Ocké M, Welch A, Boeing H, Liere M *et al*. (2000). Standardization of the 24-h diet recall calibration method used in the European Prospective Investigation into Cancer and Nutrition (EPIC): general concepts and preliminary results. *Eur J Clin Nutr* **54**, 900–917.

Turrini A, D'Addezio L, Arganini C, Camilli E, Marletta L, Ireland J (2009). Using systematic food description to aggregate foods for nutritional surveys: example of the INRAN database. Poster: 3rd International EuroFIR Congress. University of Vienna (Austria), 8–10 September 2009.