LANGUAL in the European Context

Why do we need relevant food data interchange?

Foods are involved in all major human activities:
- health and science, in the context of food correlated diseases such as cancer or diabetes, or malnutrition in developing countries;
- trade and economy, with increasing international exchange of commodities;
- regulation and politics, which depend on food hygiene, agricultural policy and trade control;
- social behavior and human sciences, as food is not only a way to survive but also a source of pleasure and involved in many social or religious habits.

Unfortunately, we are not always able to understand and compare nutritional status for different countries or people, due to the fact that there is no scientific method for describing foods. Natural language is often inadequate and even misleading to those who are not closely acquainted with the local language and culture. The situation is further confused by homonyms, synonyms, identical brand names for different products, and culinary or technological terms.[1]

Significant food description is equally of increasing importance due to labelling regulations.[2] Since September 24, 1990, nutritional labelling is regulated at the European level, following similar legislation in the USA and Canada.[3] A normalized method of describing foods will be necessary, as the obligation of nutritional labelling will have an effect on international trade. Thus, there is a veritable need for an international food language.

In this context, the CODATA (Committee on Data for Science and Technology) task group on "Systematic Nomenclature for Foods in Numeric Data Banks" was created in order to : provide an international system by which foods can be described for databases and especially for nutritional databases; design procedures for collecting information on food composition and data interchange; standardize methods of data processing to obtain certified values from aggregated data; establish guidelines for database structure. On the European level, the FLAIR (Food-Linked Agricultural and Industrial Research Concerted Action Program) - ENFANT (European Network on Food and Nutrition Tables) working party on "Food Coding Systems and Food Composition Data Bases" was created to develop a European uniform coding system in order to exchange food composition data efficiently and facilitate epidemiological surveys.

Basic principles of LANGUAL

LANGUAL (*Langua Alimentaria*) is the only generally recognized method in common use for describing, capturing and retrieving data about food. LANGUAL is an international descriptive coding system for foods, based on the principle of a faceted thesaurus, where each food coded is described by a set of standard terms grouped in facets.[4] It is a systematic method for describing foods adequately in such a manner that non-food chemists can use a database. Initially, this system was developed by the U.S. Food and Drug Administration (FDA), in cooperation with the U.S.

---

[1] W. Polacchi, "Standardized food terminology: An essential element for preparing and using food consumption data on an international basis", *Food and Nutrition Bulletin*, Vol. 8, No. 2, pp. 67-68.

[2] Commission du *Codex Alimentarius*, "Rapport de la vingtième session du comité du codex sur l'étiquetage des denrées alimentaires", Ottawa, 3-7 avril 1989.

[3] *EEC Regulation* 90/496 of 24 September 1990.

[4] H. Haendler, "Synthetic description systems for accurate data identification and selection: Principles and methods of nutritional data banks", *Int. Classif.*, 1988, vol.15n 2, 64-68.

National Cancer Institute (NCI), to provide a standardized language for describing foods and specifically for classifying food products for information retrieval.[5,6] Designed to serve their needs, each of its facets represents a subset of characteristics which specify the nutritional and/or hygienic quality of an aliment (Table 1), as for example the biological origin, the methods of cooking and conservation, and technological treatments. LANGUAL permits searches to be done on one food, on a combination of foods, or on a combination of factors within a food by computer, in an easy, efficient manner.

LANGUAL is a food description language but not a software program. It is being used within the FDA on an IBM main frame computer; a personal computer version is also being developed.

Some examples to make this clear

Table 2 presents the example of a standardized description of a food using LANGUAL: French bread produced in France according to French regulations. An interesting feature of LANGUAL is that the coded descriptions of American, English and German breads are significantly different from one another because these are not the same foods, either from a nutritional, technological or legal standpoint.[7]

Figure 1 shows, in the form of an entity-relationship diagram, how LANGUAL is used for describing foods in a data base. A typical description consists in a list of 11 to 23 standardized descriptors or terms, put together in the same thesaurus. LANGUAL contains more than 2,900 such standardized descriptors. Each descriptor belongs to a facet, specific to a set of characteristics useful for food identification, identified by an uppercase letter. In order to avoid imprecision and errors of translation, each descriptor has a scope note which contains a legal or technical definition. In order to enlarge or narrow retrieval equations, descriptors are structured as hierarchical trees within each facet, from broader to narrower terms (Table 3). For example, one might search specifically for "soft cheese" or more broadly for "cured cheese" or most broadly for cheese or cheese products, or one might aggregate all foods having "cow" or "curd" as source. The hierarchical arrangement also displays the vocabulary in a logical way to facilitate indexing and retrieval.

Often terms employed for a descriptor are very closely associated with regulation, food supply and/or culture. For example, the definition of "skimmed milk" differs from one country to another as shown in Table 4. Therefore, when precise information is available, descriptors must include a precise spectrum value instead of a simple vernacular term. In fact, a food could be coded differently in different countries or even in the same country, the goal being to obtain as close agreement as possible. This necessitates well-defined coding rules, international trials and communication among countries. The definitions of the descriptors should also be brought to a more general, international level.

How LANGUAL is used

Although it is not the only food classification and description language, LANGUAL is considered to be the most definitive at the present time. LANGUAL is translated and usable in four languages (English, French, Danish and Hungarian); two other languages should be available in 1992 (Spanish and German). It is implemented on several computers for retrieving data in about 10 databases

---

[5] A. Mc Cann et al., "FDA's Factored Food Vocabulary for food product description", *Perspectives in Practice*, March 1988, v. 88 n° 3, 336-341.

[6] E.C. Smith, "Update on Factored Food Vocabulary: LANGUAL", *Fourteenth National Nutrient Databank Conference*, 1989, University of Iowa.

[7] M. Feinberg, J. Ireland-Ripert, J.C. Favier, "LANGUAL: un langage international pour la description structurée des aliments", *Science des Aliments*, 11(1991), 193-214.

concerning chemical analysis, residue values of pesticides, toxic elements, nutrient elements, industrial chemicals, food additives and regulatory information:

- FDA Total Diet Study (quarterly analysis of typical market basket, on residue values of pesticides, toxic elements, nutrient elements and chemicals);[8]
- Total Diet Market Basket Survey;
- Food Labelling and Product Surveillance Files (FLAPS);
- Scientific Information Retrieval and Exchange Network (SIREN, food additives and regulatory information);
- USDA Nutrient Database for Standard Reference (Handbook #8);[9]
- NCI Food Component Research Database;
- Codex Alimentarius;
- National Food Consumption Survey;
- Carotenoid Foods;
- Greek foods;
- Canada: Health and Welfare Department database;
- Denmark: National Food Agency database;
- France: CIQUAL (Centre Informatique sur la Qualité des Aliments) nutritional database (REGAL);
- CREDOC (Centre de Recherche pour l'Etude et l'Observation des Conditions de vie) total diet studies.
- Hungary : food database.

Altogether, over 30,000 food products have been coded in various countries using this system.

LANGUAL adapted to specific uses

LANGUAL is used in two main ways. The first is a system of descriptors for **food databases**, which has already been implemented within the USA, Canada, France, Denmark and Hungary. The second purpose would be to link LANGUAL to a software program for dietary data collection from individuals in **epidemiological studies**.

Nutritional databases

Individuals working on national food consumption databases are interested in using LANGUAL since it is a tool for accessing other data banks and for collaboration with other countries. This language is currently being proposed as an option for an international interface standard for food databases.

To illustrate the use of LANGUAL in nutritional databases, Table 5 lists the descriptors currently used in the French REGAL database. Nevertheless, all the descriptors are potentially applicable in the European context, with the exception of US cheese classifications in table A, some native American and Asian plants or animals in table B, and specifically American meat cuts and qualities in table Z.

Epidemiological surveys

LANGUAL is a multifaceted system which could be appropriate for use in the database of the European Prospective Study on Nutrition, Cancer and Health, as description of the food consumed by the subject is required for epidemiological studies.

The CREDOC has, to date, coded 7000 food items using the LANGUAL system for use in its national consumption studies.

---

[8] J.A. Pennington, "Total diet study and Factored Food Vocabulary: LANGUAL", *Fourteenth National Nutrient Databank Conference*, 1989, University of Iowa.

[9] R. Butrum, J. Pennington, "Technology systems used for food composition data bases", in P.S. Glaeser (ed): *Computer Handling and Dissemination of Data*. Amsterdam, Elsevier Sciences Pub., 1987, p 404 ff.

However, LANGUAL may have to be modified or adapted for use in recall questionnaires. The International Agency for Research on Cancer (IARC) has drawn up a provisional list of facets (Table 6) that could be used in the European Prospective Study on Nutrition: IARC Main Food Groups and LANGUAL tables B, C, H, (E), G(F), K, Z. The IARC will, moreover, access the LANGUAL descriptors and facets to determine whether they are adequate for use in Europe for nutritional epidemiological studies. They have already suggested improvements that can be made to the LANGUAL thesaurus; for example, table G (cooking method) could be expanded to include cooking times and temperatures.

Working from a different point of view, that of the use of the LANGUAL descriptors in coding foods, the CIQUAL proposed an adaptation of the LANGUAL hierarchy for the IARC main food groups (Annex).

The Minnesota Nutrition Data System, which contains a hierarchical order of foods and a multifaceted data description and collection methodology, could be adapted to include LANGUAL. A working group will be established between the US FDA and the Nutrition Coordinating Centre of the University of Minnesota to assess whether the NDS can be mapped with LANGUAL. This will require deciding how many factors are not already in NDS and how many should be added. Another working group at the IARC will evaluate the NDS system and examine how and whether it could be adapted for use in Europe. The results of these studies should help adapt LANGUAL to the collection of consumption data in Europe.

Feasibility of LANGUAL in the ENFANT-EUROFOODS project

Evaluation

The evaluation of the LANGUAL faceted descriptor system in the European context that has been done so far consists in its adaptation to **nutritional databases** of countries (France, Denmark, Hungary) outside the USA.

A formal test of LANGUAL for use in **nutritional studies** has not yet been undertaken. However, at the CREDOC, an evaluation of LANGUAL is being undertaken to determine whether food items from consumption surveys can be properly described without ambiguity; it appears from their work that a descriptor-based structure is the only way to handle a food database.

It has been proposed to test the LANGUAL system for epidemiological surveys, where a faceted descriptor system would be valuable for describing foods consumed and accessing foreign data banks for missing values. A test must be designed, a list of criteria be drawn up and methods evaluated through a random selection of 24-hr recalls to be coded in each country. In this context, the LANGUAL system may be modified and simplified to suit the needs of epidemiological studies, and possibly integrated to a coding system (EUROCODE), with a method of linking coding/descriptor systems with existing food composition tables.

Problems raised by the use of LANGUAL in the European context:

- Translations and adaptations

The LANGUAL thesaurus already exists in English, French, Danish and Hungarian, and is in the process of being translated into Spanish and German. Priority will be given to the translation of LANGUAL to the languages of the other countries taking part in the European Prospective Study on Nutrition: Italy and the Netherlands. Until now, all translations have been carried out by the users of the LANGUAL system. To accelerate the process, additional funding could be requested from the EEC and CODATA; the translations can be maintained (new terms, modifications,...) by national LANGUAL committees (see below).

- Organizational needs

Until the present time, LANGUAL has been centered in the USA, and decisions have been taken there. A committee of experts meets once a month at FDA to discuss all changes and additions to LANGUAL which arise from work on the LANGUAL dictionary in any center. More time is now being spent on applying LANGUAL and less time on establishing the system and the theoretical basis of the language. Changes have been made to the original US version of LANGUAL because of suggested modifications arising from the adaptation of LANGUAL for use in data banks of other countries.

In the future, a more international administrative hierarchy will be necessary for the maintenance and updating of the system over time (Figure 2).
- The STEERING COMMITTEE is presently headed by Bradley Rosenthal and Ritva Butrum in the USA and Max Feinberg in Europe.
- USER GROUPS (users of data banks): dietitians, consumers, industry.
- TECHNICAL COMMITTEE: parallel technical committees will exist in Washington and in Europe, with exchanges of information. The LANGUAL working group will continue to meet in Washington once a month, with a European representative every 3-4 months; similar meetings will be held in Europe. Minutes will be written and exchanged between committees.
- TELECOMMUNICATIONS AND COMPUTERS: possibilities of international exchange of information by BITNET; an inventory of current hard and software will be made.
- DATA/STATISTICS (numerical composition): methods of gathering, aggregating and imputing of data.
- COMMUNICATIONS/PUBLICITY: multilingual papers published in the LanguaLine newsletter.

There may be evolution of the language over time, with the appearance of new food products and new needs. For the sake of coherence, the system should be developed jointly within Europe. It is also possible to add version numbers of LANGUAL to copies of the language which are released so that retroactive work can be done.

- Competences required

The European TECHNICAL COMMITTEE will be composed of members from the different European countries using the LANGUAL descriptive system; they will represent their national committees. These committees should include experts in food databases, food technology, nutrition and statistics.

- Training

Training sessions must be organized in Europe; courses can be given by persons who are currently using LANGUAL and who have already followed a LANGUAL training course. Some funding can be obtained from the US FDA for training courses in Washington; other funds could possibly be obtained at the EEC level. Some training materials already exist in English (LANGUAL User's Manual); others should be created for specific national contexts. Use of a demonstration program, like that of the NCI, would be useful in this context.

Figure 1. Use of LANGUAL in a nutritional database

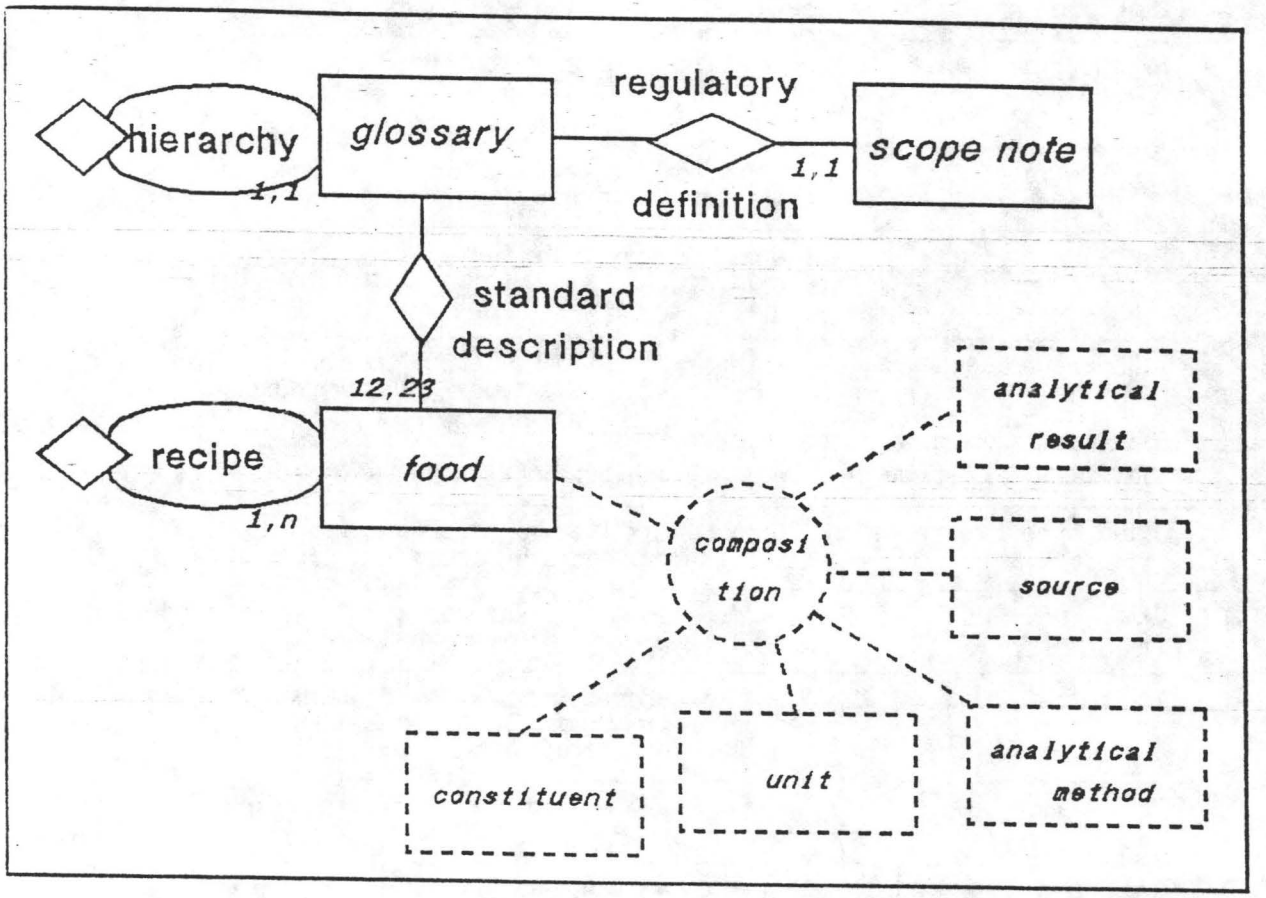

Figure 2. LANGUAL administrative hierarchy

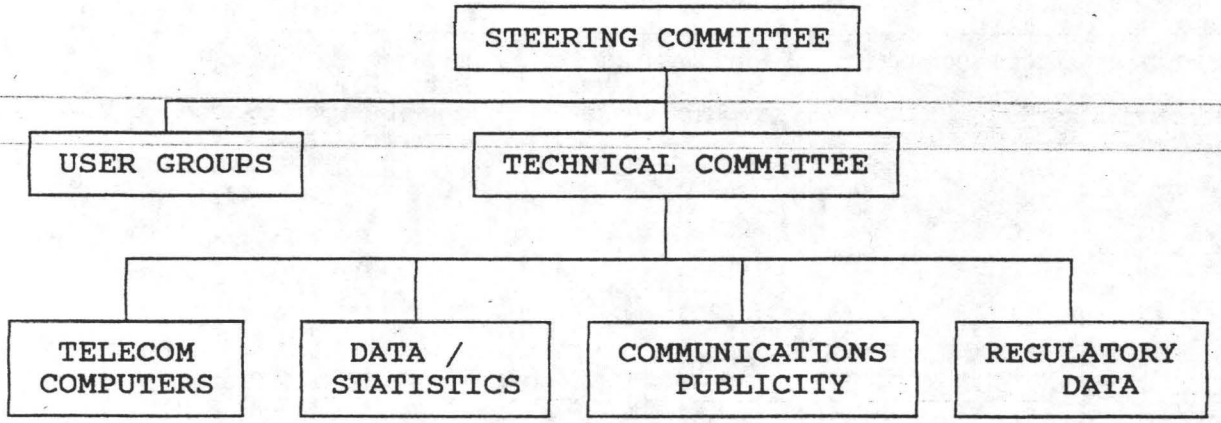

Table 1. LANGUAL facets

Facet	Definition	
A	Product type	Family or group of foods defined by common consumption, functional or manufacturing characteristics
B	Food Source	Animal, plant or chemical source form which the product or the primary ingredient is derived
C	Par of plant/animal	Anatomical part of the plant or animal from which the food product or its major ingredient is derived (meat, milk, root, sugar)
Z	Adjunct characteristics	Quality criteria (label, meat cut, plant maturity) and other characteristics (type of crust, casing, beverage mix)
E	Physical state or shape	Physical state of food product as a whole (solid, liquid)
F	Extent of heat treatment	Extent the food has been modified in processing by the application of heat (raw, cooked)
G	Cooking method	Process by which a food product is cooked (broiled or grilled, deep-fried, cooked with steam)
H	Treatment applied	All physical or chemical treatments applied to the product or its major ingredients; also describes additives and ingredients
J	Preservation	Primary method used to prevent microbial and enzymatic spoilage
K	Packing medium	Substance in which the food is packed for preservation and handling and/or palatability
M	Container or wrapping	Defines the main container material, the container form and the liner, lids and ends material
N	Food contact surface	Material or materials which actually touch the food product
P	Consumer/dietary group	Group for which the food product is marketed (regular diet, low fat)
Geographic places and Regions		Identifies places of origin, transformation and consumption of food product (ISO national codes implemented to include regions and fishing zones)

Table 2. LANGUAL codes for the bread produced in France

Factor		Code	Descriptor
A	Product type	0178	Bread
B	Food source	1421	Soft wheat (Triticum aestivum)
C	Part of plant/animal	0208	Seed or kernel, skin removed, germ removed
E	Physical state or shape	0105	Whole shape achieved by forming, thick. 1.5-7 cm
F	Extent of heat treatment	0003	Complete heat transformation
G	Cooking method	0005	Baked or roasted
H	Treatment applied	0256	Carbohydrate fermented
J	Preservation	0003	No preservation method used
K	Packing medium	0003	No packing medium used
P	Consumer/dietary group	0024	Human food, no age specification, regular diet
M	Container or wrapping	0003	No container or wrapping used
N	Food contact surface	0003	No food contact surface present

Table 3. Excerpts from the LANGUAL thesaurus

A) Product type
- *Dairy product*
  - Cheese or cheese product
    - Natural cheese
      - Cured cheese
        - Hard grating cheese
        - Hard cheese
        - Semi-soft cheese
        - Soft cheese
      - Uncured cheese

B) Food source
- *Animal used as food source*
  - Meat animal
    - Cattle
      - Cow

C) Part of plant or animal
- *Part of animal*
  - Milk or milk component
    - Cream or cream component
    - Curd
    - Milk
    - Whey

Table 4. Example of different food regulations. The fat content of milk

Food name	Europe [10]	U.S.A. [11]
Whole milk	≥ 3.5 % fat	≥ 3.25 % fat
Half-Skimmed or Low fat	1.5 to 1.8 % fat	0.5 %, 1 %, 1.5 % and 2 % fat
Skimmed	≤ 0.3 % fat	≤ 0.5 % fat

[10] *EEC Regulation* 1411/71 of 29 June 1971.

[11] *Code of Federal Regulations* 21, chapter 1.

Table 5. Number of occurrences of the LANGUAL descriptors in the French nutrient database

	Facet A	#
A100	Pie, unsweetened, or pizza	1
A101	Cultured milk product	34
A106	Prepared grain or starch product	6
A110	Pasteurized process cheese	3
A117	Pasteurized cheese product	1
A125	Grain or starch product	19
A129	Refined or partially-refined food product	94
A140	Pie crust, sweetened	8
A148	Milk or milk product	53
A149	Milled grain or starch product	37
A151	Sweet roll or bun	45
A152	Vegetable or vegetable product	5
A160	Pie crust, unsweetened	16
A178	Bread	35
A183	Quick bread, sweetened	1
A185	Uncured cheese	19
A191	Bakery product	4
A194	Pancake or waffle	13
A202	Pastry, sweetened	87
A203	Cookie	83
A210	Cake	101
A214	Glaze	4
A221	Sausage or luncheon meat	358
A222	Dessert	10
A228	Snack food	3
A242	Cracker	2
A248	Doughnut	1
A253	Pie, sweetened	1
A261	Egg or egg product	1
A269	Icing or frosting	7
A275	Macaroni or noodle product	27
A279	Cured meat	80
A310	Extra hard cheese (Codex)	1
A311	Hard cheese (Codex)	4
A313	Semi-soft cheese (Codex)	19
A314	Soft cheese (Codex)	36

	Facet B	#
B0001	Food source not known	3
B1012	Sugar-producing plant	52
B1017	Oil-producing plant	5
B1052	Sour cherry	1
B1064	Huckleberry	1
B1079	Durum wheat	41
B1105	Cattle and swine	24
B1129	Trout family	1
B1134	Meat animal	12
B1136	Swine	289
B1161	Cattle	16
B1183	Sheep	3
B1192	Black currant	1
B1201	Cow	190
B1207	Prune	2
B1209	Date	1
B1210	Cotton	1
B1218	Potato	2
B1219	Oat	2
B1222	Fish	8
B1226	Sesame	1
B1228	Common carp	1
B1229	Horse	2
B1230	Barley	6
B1232	Corn	2
B1236	Turkey	3
B1245	Apple	7
B1248	Rape	5
B1249	Papaya	1
B1253	Goose	17
B1272	Almond	3
B1276	Tomato	1
B1283	Melon	2
B1285	Millet	1
B1286	Palm	7
B1290	Walnut	1
B1297	Animal used as food source	8
B1298	Quince	1
B1299	Olive	3
B1306	Cherry	1
B1312	Wheat	21
B1313	Rye	12
B1316	Duck	14
B1318	Cacao	4
B1322	Rice	5
B1323	Rabbit	4
B1324	Grain	2
B1329	Grape	3
B1337	Peanut	5
B1339	Orange	2
B1344	Pear	2
B1347	Plant used as food source	1
B1349	Calf	10
B1350	Peach	1
B1358	Triticale	1
B1364	Rhubarb	2
B1378	Sugar cane	2
B1379	Field corn	8
B1380	Vegetable corn	5
B1393	Strawberry	2
B1394	Gooseberry	1
B1396	Menhaden	1
B1405	Buckwheat	2
B1414	Herring	1
B1416	Pistachio	1
B1421	Soft wheat	252
B1448	Sorghum	2
B1452	Soybean	1
B1454	Popcorn	2
B1456	Pumpkin	1
B1457	Chicken	1
B1470	Avocado	1
B1515	Poppy	1
B1529	Apricot	2
B1530	Kumquat	1
B1533	Filbert	1
B1536	Coconut	4
B1544	Chestnut	1
B1547	Sunflower	13
B1548	Safflower	3
B1558	Sardine	1
B1563	Poultry or game bird	19
B1580	Field corn, yellow	1
B1583	Deer	1
B1584	European chestnut	1
B1667	Sheanut	1
B1691	Roselle	1
B1700	Flax	1
B1701	Lupine bean	1
B1703	Castor bean	1
B1707	Babassu palm	1
B1713	Hen	28
B1735	Borage	1
B1842	Atlantic cod	1
B2038	Grape, corinth	2
B2109	Evening primrose	1
B2110	Illipe	1
B2242	Swine and calf	10
B2243	Cattle and swine and sheep	1
B2244	Cattle and sheep	1
B2248	Boar	2
B2610	Ewe	2
B2611	Doe (goat)	12

	Facet C	#
C001	Part of plant or animal not known	3
C101	Stem or stalk (without leaves)	2
C102	Germinated or sprouted seed	2
C103	Meat part	8
C108	Sugar	4
C111	Skin or bran (pericarp)	11
C112	Pod with seeds removed	1
C113	Milk or milk component	2
C119	Seed or kernel, skin present, germ removed	3
C125	Skeletal meat part, without bone or shell	35
C129	Organ meat, cardiovascular system	1
C132	Seed or kernel, skin removed, germ present	2
C133	Seed or kernel, skin present, germ present	23
C134	Seed or kernel, skin removed	3
C135	Seed or kernel, skin present	1
C138	Fruit or berry, peel removed	1

C139	Fruit or berry, peel present, core, core or seed removed	7
C140	Fruit or berry, peel present, core, core or seed present	5
C142	Germ and bran	3
C148	Root, stem, leaf or flower	1
C154	Whipping cream	3
C155	Seed or kernel	10
C160	Light cream	2
C161	Heavy cream	1
C167	Fruit or berry (anatomical part)	1
C175	Skeletal meat part	4
C176	Liver	48
C179	Butter	3
C185	Blood	21
C189	Tongue	5
C190	Fat or oil	92
C192	Tripe	9
C193	Skin, animal	3
C195	Cream	1
C208	Seed or kernel, skin removed, germ removed (endosperm)	305
C210	Sucrose	49
C212	Intestine	13
C218	Organ meat	1
C225	Whole egg without shell	22
C228	Extract, concentrate or isolate of plant or animal	2
C229	Fruit or berry, peel removed, core, core or seed removed	15
C235	Milk	106
C240	Root, tuber or bulb, without peel	2
C241	Egg white, albumen	1
C242	Seed on cob, with or without husk	1
C245	Curd	86
C253	Fat, trim	9
C267	Skeletal meat part, without bone, with skin	27
C268	Skeletal meat part, without bone, without skin	98
C269	Skeletal meat part, without bone and skin, with separable fat	111
C270	Skeletal meat part, without bone and skin, without separable fat	44
C274	Whole egg	5

Facet E		#
E001	Physical state, shape or form not known	6
E101	Medium ground and sifted (bolted)	1
E102	Liquid, high viscosity	8
E105	Whole, shape achieved by forming, thickness 1.5-7 cm.	349
E106	Finely ground	23
E110	Semiliquid with solid pieces	17
E111	Cut into short pieces	1
E117	Medium ground	19
E119	Semisolid with smooth consistency	113
E123	Liquid, low viscosity, with no visible particles	27
E124	Sliced, thick, between 0.5 and 1.5 cm.	18
E131	Whole	43
E134	Semisolid with solid pieces	40
E135	Semiliquid with smooth consistency	6
E139	Liquid, high viscosity, with no visible particles	74
E140	Whole, shape achieved by forming, thickness 0.3-1.5 cm.	71
E144	Semisolid	9
E146	Whole, shape achieved by forming, thickness >7 cm.	188
E147	Whole, shape achieved by forming	74
E150	Whole, natural shape	59
E151	Solid	25
E153	Whole, shape achieved by forming, thickness <0.3 cm.	46

Facet F		#
F01	Extent of heat treatment not known	99
F03	Not heat treated	223
F14	Fully heat treated	752
F18	Partially heat treated	145

Facet G		#
G001	Cooking method not known	96
G003	Cooking method not applicable	498
G004	Cooked by dry heat	31
G005	Baked or roasted	364
G006	Broiled or grilled	4
G008	Griddled	5
G009	Popped	1
G010	Toasted	9
G012	Cooked by moist heat	34
G013	Cooked in water or water-based liquid	108
G014	Boiled	2
G015	Boiled and drained	10
G016	Boiled in large amount of liquid	6
G017	Boiled in small amount of liquid	9
G018	Boiled and undrained	1
G019	Braised	3
G020	Simmered, poached or stewed	4
G023	Steamed without pressure	1
G025	Cooked with added fat or oil	42
G026	Cooked in small amount of fat or oil	58
G029	Deep-fried	1
G031	Cooked in container immersed in water or steam	25

Facet H		#
H001	Treatment applied not known	26
H003	No treatment applied	31
H100	Flavoring or spice extract or concentrate added	34
H101	Lactic acid fermented	53
H107	Lactic acid-other organism fermented	63
H110	Smoked by adding smoke concentrate	2
H117	Flavoring or taste ingredient added	64
H118	Smoked by smoke infiltration	21
H122	Jelly, jam or preserve added	19
H132	Pudding or custard added	1
H133	Certified color added	8
H134	Protein removed	4
H136	Sugar or sugar syrup added	138
H138	Water removed	123
H143	Cheese added	2
H146	Starch added	27
H147	Fruit added	76
H148	Water added	4
H149	Honey added	12
H150	Color added	9
H151	Spice added	8
H152	Grain added	159
H154	Dextrose added	5
H155	Maltose added	27
H156	Molasses added	3
H158	Sucrose added	209
H160	Alcoholated	78
H161	Fat removed	12
H162	Calcium added	5
H163	Vitamin added	3
H165	Gelatin added	63
H166	Mushroom added	11
H169	Instantized	1
H171	Poultry added	13
H172	Smoked or smoke-flavored	13
H173	Salted	47
H174	Hydrogenated	9
H175	Carbonated	1
H177	Nut or seed added	73
H178	Aerated	23
H180	Food added	1
H181	Iron added	1
H182	Candied	8
H184	Milk added	235
H185	Egg yolk added	68
H186	Egg added	283
H188	Breaded or batter-coated	8
H191	Meat added	83
H192	Textured	1
H193	No salt added	9
H194	Nutrient or dietary substance added	1
H197	Bleached	43
H198	Debittered	1
H200	Acidified	2
H205	Egg white added	34
H206	Alkalized	47
H207	Filled or stuffed	79
H211	Invert sugar added	1
H212	Vegetable added	54
H213	Vitamin A or carotenes added	5
H216	Vitamin B added	1
H221	Fat or oil added	150
H225	Ingredient added	3
H227	Flavoring, spice or herb added	73
H229	Flavoring, spice or herb added, natural	363
H231	Chocolate or cocoa added	52
H232	Alcohol fermented	2
H233	Fat or oil coated	7
H241	Previously frozen	8
H242	Dairy product added	2
H247	Fat partially removed	35
H248	Fat fully removed	16
H249	No nitrite/nitrate added	1
H252	Enzymatically modified	4
H253	Cured or aged	93

H256	Carbohydrate fermented	89
H258	Unpurified, unrefined	15
H259	Rehydrated	9
H260	Whey added	5
H262	Animal fat or oil added	272
H263	Vegetable fat or oil added	47
H271	Butter added	154
H272	Margarine added	20
H275	Liquid injected	3
H280	Sugar syrup or sugar syrup solids added	4
H288	Cured or aged less than 2 weeks	5
H289	Cured or aged two weeks to 1 month	26
H290	Cured or aged 1 to 2 months	5
H291	Cured or aged 2 to 4 months	10
H292	Cured or aged 4 to 6 months	3
H293	Cured or aged 6 to 12 months	2
H294	Cured or aged 12 months or over	1
H296	Cream added	30
H297	Milk or milk product added	10
H298	Clotting agent added	90
H303	Sorbitol added	6
H306	Homogenized or emulsified	56
H307	Organ meat added	53
H308	Poultry organ meat added	2
H319	Wheat added	4
H331	Interior mold cured	2
H351	Chemical leavening agent added	99
H353	Glazed, iced or coated	29
H360	Caramalized sugar added	7
H364	Dried or candied fruit added	32
H366	Flattened	39

	Facet J	#
J001	Preservation method not known	175
J003	No preservation method used	367
J100	Preserved by adding chemicals	10
J102	Sterilized before filling	3
J106	Preserved by smoking	3
J108	Preserved by treatment with chemicals	41
J116	Dehydrated or dried	94
J117	Heat dried	1
J120	Preserved by thermal processing	57
J123	Sterilized by heat, canned	20
J131	Refrigerated	361
J133	Spray-dried	3
J135	Pasteurized by heat	50
J136	Frozen	7
J141	Naturally dried	1
J144	Artificially heat dried	14
J145	Preserved by reducing water activity	8
J147	Sterilized at ultra high temperature (uht)	3

	Facet K	#
K001	Packing medium not known	9
K003	No packing medium used	1160
K004	Packing medium, other	4
K015	Packed with aerosol propellant	1
K026	Packed in fat or oil	2
K034	Packed in gravy or sauce	12
K035	Packed in gelatin	4
K042	Packed in broth	8
K046	Packed in animal fat or oil	16

	Facet P	#
P001	Consumer group/dietary use not known	18
P003	Consumer group/dietary use not applicable	10
P024	Human food, no age specification, regular diet	1153
P037	Human food, reduced sodium	4
P038	Human food, low sodium	2
P039	Human food, low fat	12
P040	Human food, reduced fat	32
P047	Human food, sodium special diet	3
P048	Human food, dietary fiber claim	3

	Facet Z	#
Z004	Light meat	6
Z040	Jowl, trimmed	17
Z041	Foot	3
Z047	Cut of meat, composite	4
Z051	Unripe or immature	1
Z072	Cheese rind, washed or brushed	12
Z073	Cheese rind, moldy	18
Z087	Controlled name	21
Z093	Inedible sausage casing	3
Z094	Edible sausage casing	99
Z095	Sausage casing not known	54
Z096	No sausage casing	26
Z098	Solid containing very small particles (< 1 mm)	65
Z099	Solid containing small particles (1-4 mm)	27
Z100	Solid containing medium or large chunks (8 mm)	142
Z125	Flanchet	5
Z127	Plate-de-cote	2
Z129	Collier	1
Z130	Basses-cotes	1
Z131	Raquette	12
Z136	Cuisse	37
Z141	Aloyau	3
Z142	Entre-cote	2
Z147	Cheese rind, natural	19

approx. 420 LANGUAL descriptors presently used in REGAL data bank

Table 6. Adaptation of the LANGUAL Hierarchy to IARC epidemiological surveys

A. Main Food Groups (IARC Provisional classification system)

01. Cereals and cereal products (A125)
02. Vegetables (A152)
03. Potatoes
04. Fruits (A143, 306)
05. Meat and meat products (A150, 273)
06. Fish, crustaceans, molluscs (A267)
07. Eggs (A261)
08. Dairy Products (A164)
09. Fats, oils (A129)
10. Beverages (A229)
11. Miscellaneous

B. Food Source

Animal used as Food Source
Amphibian (B1624)
Fish or lower water animal (B1021)
Insect (B1220)
Meat animal (B1134)
Poultry or game bird (B1563)
Plant used as Food Source
Fruit-producing plant (B1140)
Grain or seed producing plant (B1047)
Plant used for producing extract or concentrate (B1013)
Vegetable-producing plant (B1579
Food Source not known

C. Part of Plant or Animal

Extract, concentrate of isolate of plant or animal
Carbohydrate or related compound (C280)
Essential oil (C260)
Fat or oil (C190)
Multicomponent extract (C159)
Protein extract (C236)
Part of Animal
Animal body or body part (127)
Egg (C194)
Milk or milk component (C113)
Part of plant
Fruit of seed (C165)
Root, stem, leaf or flower (C148)

E. Physical state, shape or form (only required for a few food items)

Liquid (E130)
Semiliquid (E103)
Semisolid (E144)
Solid (E151)
Multiple (E108)

G. Cooking method

Cooked by dry heat
Baked or roasted (G005)
Broiled or grilled (G006)
Griddled (G008)
Popped (G009)
Toasted (G010)
Cooked by microwave (G011)
Cooked by moist heat
Cooked in steam (G021)
Cooked in water or water-based liquid (G013)
Cooked with fat or oil
Cooked with added fat or oil (G025)
Cooked with inherent fat or oil (G030)
Method of heating container
Cooked in container immersed in water or steam (G031)
Cooked in double boiler (G033)
Cooked in water bath (G034)

H. Treatment applied

Component removed

Alcohol removed (H285)
Carbohydrate removed (H266)
Fat removed (H161)
Protein removed (H134)

Component substituted

Fat substituted (H208)
Protein substituted (H250)

Food modified

Instantized (H169)

Ingredient added

Carbohydrate or related compound added (H301)
Fat or oil added (H221)
Filled or stuffed (H207)
Flavouring or taste ingredient added (H117)
Food added (H180)

Water added or removed

Water added (H148)
Water removed (H138)

K. Packing medium

Packed in edible medium

Packed in broth (K42)
Packed in cream or milk (K43)
Packed in fat or oil (K26)
Packed in fruit juice (K39)
Packed in gelatlin (K35)
Packed in gravy or sauce (K34)
Packed in ink (K44)
Packed in salt brine (K18)
Packed in sweetened liquid (K23)
Packed in vegetable juice (K16)
Packed in vinegar (K29)
Packed in water (K17)

Z. Adjunct characteristics of food

Adjunct characteristics of meat, poultry or fish

Cut of meat (Z146)
Extent of fat trim (Z107)

Formulated mix (Z074)

Preparation establishment

Commissary prepared (Z113)
Food industry prepared (Z112)
Home prepared (Z109)
Restaurant or fast food prepared (Z119)
Street vendor prepared (Z120)