

# LANGUAL : UN LANGAGE INTERNATIONALE POUR LA DESCRIPTION STRUCTURÉE DES ALIMENTS

LANGUAL : AN INTERNATIONAL LANGUAGE  
FOR THE STRUCTURED DESCRIPTION OF FOODS (\*)

---

M. FEINBERG (1), J. IRELAND-RIPERT (1), J.-C. FAVIER (2)

---

## RÉSUMÉ

Cet article traite de l'organisation des nouvelles banques de données sur la composition des aliments en s'appuyant sur l'exemple du Répertoire Général des Aliments (REGAL), la banque de données française développée par le Centre Informatique sur la Qualité des Aliments (CIQUAL).

D'abord, on y évoque la codification des nutriments et les principes de standardisation des formules de calcul de certains nutriments. Mais surtout on y propose une solution pour l'identification précise des aliments, outil indispensable pour assurer des transferts de données entre banques de différents pays qui soient cohérents. Son principe est d'avoir, à côté d'une dénomination usuelle, une description claire, complète, scientifique et qui se prête à des vérifications automatisées de similitude.

Un tel système de codification descriptive des aliments est en cours d'élaboration entre plusieurs pays. Le système présenté s'appelle LANGUAL (Langua Alimentaria). Basé sur le principe d'un thesaurus en facettes, chaque aliment à coder est décrit par un ensemble de termes standardisés, eux-mêmes regroupés en facettes. Chaque facette représente un sous-ensemble de caractéristiques, comme l'origine biologique, les méthodes de conservation et de cuisson ou les traitements technologiques, qui spécifient la qualité nutritionnelle et/ou hygiénique d'un aliment.

**Mots clés :** *nutrition, banque de données, composition des aliments, codification.*

---

(1) Institut National Agronomique Paris-Grignon, Centre Informatique sur la Qualité des Aliments, 16 rue Claude-Bernard, 75231 Paris cedex 05, France.

(2) Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM).

(\*) Version anglaise disponible auprès de M. FEINBERG.

ORSTOM Fonds Documentaire

29 OCT. 1993

N° 38 297 ep1

Cote B

## SUMMARY

The article concerns the organization of a food databank, following the example of the Répertoire Général des Aliments (REGAL), the French food databank developed by the Centre Informatique sur la Qualité des Aliments (CIQUAL).

The coding of nutrients is first discussed, and the system for formula translation used to calculate certain nutrients in databanks is exposed. Above all, a solution for the precise identification of foods is proposed, essential to ensuring consistency of information interchange between databanks of different countries. Food names must be completed with clear, complete and scientific descriptions, allowing systematic verification of similitude.

Such a descriptive coding system for foods is being developed in several countries. The system, LANGUAL (Langua Alimentaria), is based on the principle of a faceted thesaurus, where each food coded is described by a set of standard terms grouped in facets. Each facet represents a subset of characteristics which specify the nutritional and/or hygienic quality of an aliment, as for example the biological origin, the methods of cooking and conservation, and technological treatments.

**Key-words:** nutrition, databank, food composition, coding system.

---

## 1 - INTRODUCTION

---

Depuis une dizaine d'années on constate, dans de nombreux pays, un regain d'intérêt pour la chimie des aliments qui se manifeste par la création de banques de données sur la composition des aliments. Récemment même, cette tendance a débouché sur le lancement du *Journal of Food Composition and Analysis* qui publie les travaux spécifiques de ce domaine situé à l'interface de la chimie analytique et de l'informatique.

On peut en partie expliquer ces activités du fait de l'obsolescence des données contenues dans les tables de composition. En effet, les premières éditions de ces ouvrages de référence remontent aux années 1930. Les nombreuses mises-à-jour intervenues depuis n'ont pas toujours été réalisées en profondeur ni en fonction des nouveaux aliments et des nouvelles méthodes d'analyse. Mais cet argument est insuffisant, car les progrès de la nutrition ne justifient pas totalement le développement de banques de données coûteuses.

Les motivations vis-à-vis de la création de banques de données sur la composition des aliments sont aussi d'une autre nature : ce sont des outils aujourd'hui indispensables pour régir les rapports entre producteurs, transformateurs, distributeurs et consommateurs. En fait, alors que l'emploi de méthodes physico-chimiques d'analyse pour le contrôle de la qualité des aliments se généralise, on constate une pénurie de valeurs références fiables. En outre, les analyses circulaires organisées entre des laboratoires spécialisés prouvent à l'évidence le manque de justesse des méthodes (HOLMANN et KATAN, 1985). Les moyens nécessaires à l'amélioration de cette situation passent, bien

sûr, par un contrôle de qualité, une normalisation et une validation des méthodes d'analyse, mais aussi, par la création de valeurs de référence validées : c'est une tâche pour laquelle les banques de données peuvent être des outils de prédilection.

De fait, les utilisateurs actuels de données de composition ont des besoins variables, car ils appartiennent à des professions diverses, et évolutifs, en raison de contraintes économiques. On rencontre des nutritionnistes, des diététiciens, des industriels producteurs et/ou transformateurs, des distributeurs, des responsables de marketing ou des consommateurs.

Pour répondre à la diversité de leurs requêtes, il faut donc prévoir des modes d'accès ou de consultation très souples. Ce sont, par exemple, des livres thématiques (FEINBERG *et al.*, 1987), mais aussi des fichiers magnétiques intégrables dans des logiciels spécialisés, ou encore des moyens d'accès en-ligne et télématiques. Là encore les banques de données offrent des solutions intéressantes.

---

## 2 - L'ORGANISATION D'UNE BANQUE DE DONNÉES SUR LA COMPOSITION DES ALIMENTS

---

A titre d'illustration nous présenterons surtout le Répertoire Général des Aliments (REGAL) qui est la banque de données française développée par le Centre Informatique sur la Qualité des Aliments (CIQUAL), unité de recherche du Centre National d'Etudes Vétérinaires et Alimentaires (CNEVA).

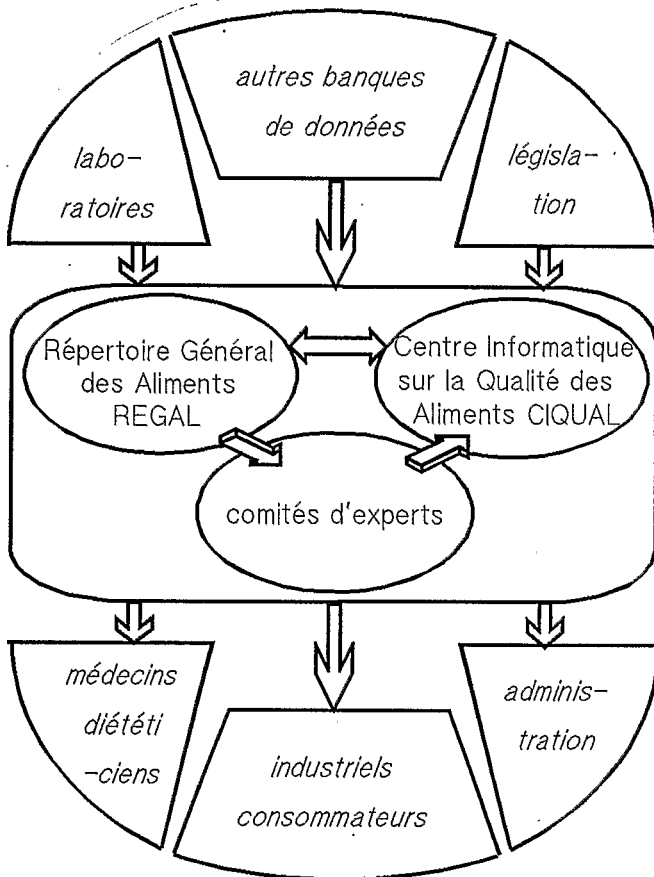
Les conditions de son fonctionnement sont très similaires à celles des banques développées par d'autres pays. Ses spécifications peuvent se comprendre à travers le schéma global des flux d'informations gérés par le CIQUAL (*fig. 1*). L'objectif est de produire, à partir de données obtenues dans des laboratoires d'analyse, des fiches signalétiques d'aliments consommés en France qui pourront servir de véritables références scientifiques, commerciales ou administratives.

Les résultats d'analyse forment donc la matière première. Pour fabriquer ces fiches, ils sont regroupés les uns aux autres à l'aide d'un Système de Gestion de Banque de Données (SGBD). Dans le cas de REGAL c'est un SGBD relationnel implanté sur mini-ordinateur, accessible par un langage d'interrogation qui suit la norme SQL (Structured Query Language). Pour environ 5 000 aliments différents, le volume final des données est estimé à 60 millions d'octets, auxquels s'ajoutent 100 millions d'octets sous forme de fichiers d'index et de programmes : ces chiffres montrent l'importance des systèmes d'indexation.

Selon le schéma général de fonctionnement qui a été adopté, les données brutes sont recueillies auprès de laboratoires d'analyse spécialisés dans l'analyse des aliments de l'homme. Elles sont mises en forme, saisies, stockées puis traitées par le CIQUAL.

Ces traitements sont définis comme une *consolidation* des données brutes. Les données consolidées sont soumises à des comités d'experts chargés de définir leur *validité*, tant du point de vue analytique que nutritionnel. Il est donc important de toujours pouvoir retrouver l'origine des données qu'elles soient brutes ou consolidées (fig. 1).

Une représentation commode de la banque de données consiste à la décrire comme un tableau à deux entrées. C'est d'ailleurs cette présentation qui est utilisée dans les tables de composition traditionnelles (RANDOIN *et al.*, 1985 ; RENAUD et ATIE, 1986 ; PAUL et SOUTHGATE, 1978 ; USDA, 1976-1986). Par exemple, les lignes contiennent les noms d'aliments, les colonnes les noms de nutriments et les cellules les valeurs de référence. Il apparaît rapidement que cette disposition en tableau engendre une perte de place importante dès que le nombre de nutriments augmente ; il y a très vite plus de données manquantes que de valeurs de référence.



**Figure 1**  
Schéma global des flux d'information

**Figure 1**  
Global diagram of information flow

Pour éviter cet inconvénient, il faut donc faire appel à une organisation plus élaborée que nous avons décrite par ailleurs (FEINBERG *et al.*, 1991). Cependant, pour assurer une gestion exacte des données, il est indispensable de pouvoir identifier correctement les aliments et les nutriments ; il faut pouvoir localiser les lignes ou les colonnes du tableau sans ambiguïtés.

C'est pourquoi, afin de mettre en place une banque de données sur la composition des aliments, il faut réunir un certain nombre de conditions préalables :

- 1 - Choisir des *systèmes de codification* des noms d'aliments et de constituants qui soient adaptés aux besoins propres d'une telle banque et harmonisés avec ceux d'autres banques pour permettre un échange de données.
- 2 - Organiser des réseaux de laboratoires pour la fourniture de données prêts à participer à des analyses inter-laboratoires garantissant la qualité des résultats.
- 3 - Fixer des algorithmes de consolidation statistique, en vue d'obtenir les valeurs de référence qui pourront être validées par des comités d'experts.

### 3 - LE SYSTÈME DE CODIFICATION DES NUTRIMENTS

Le repérage et la définition exacte des nutriments ne posent pas de problème majeur. Il est cependant utile de faire une distinction entre les constituants réellement déterminés par les méthodes physico-chimiques d'analyse et les nutriments qui correspondent aux formes assimilables des constituants. Sur le plan chimique on détermine des caroténoïdes, alors que les nutritionnistes veulent connaître l'activité vitaminique A de l'aliment.

Si la plupart des nutriments correspond effectivement à des constituants clairement identifiés sur le plan chimique, des difficultés persistent pour certains concepts qui sont encore mal précisés, comme les fibres, les glucides disponibles, etc. On peut partiellement remédier à ces ambiguïtés en associant une méthode d'analyse au nom du nutriment.

**Tableau 1**  
Système de repérage des nutriments INFOODS

**Table 1**  
*INFOODS food component tagnames*

Code	Constituant
NA	Sodium en g
FAPUN3F	Somme des acides gras poly-insaturés de la série n-3, exprimée en mg par g d'acides gras totaux
FIBSOL	Somme des constituants solubles dans l'eau, déterminés par la méthode AOAC pour la fibre alimentaire
FIBINS	Somme des constituants insolubles dans l'eau, déterminés par la méthode AOAC pour la fibre alimentaire
FIB-	Fibre, déterminée selon une méthode inconnue

**Tableau 2.1**

Produit

**Table 2.1**

Product

$$[A|C|U] = a_0 \cdot [A|C_1|U_1] \cdot [A|C_2|U_2]^{a_1} ; -1, a_1, +1$$

Protéines à partir de l'azote	(1)	$[A protéines g/kg] = k_p \cdot [A N_{total} g/kg]$
Conversion des unités d'énergie		$[A énergie kcal] = 0,239 \cdot [A énergie kJ]$
Expression des acides gras en g/kg	(1)	$[A AG g/kg] = k_{AG} \cdot [A AG \% AG_{total}] \cdot [A lipides g/kg]$
Acides aminés en mg/g d'azote		$[A AA g/kg] = 0,001 \cdot [A AA mg/gN] \cdot [A N g/kg]$
Activité vitamine A à partir du $\beta$ -carotène (1)		$[A vitamine A g/kg] = k_b \cdot [A carotène UI/kg]$

**Tableau 2.2**

Décalage

**Table 2.2**

Transposition

$$[A|C|U] = a_0 + a_1 \cdot [A|C_1|U_1]$$

Matière sèche à partir de l'humidité	(2)	$[A mat. s\grave{e}c. g/kg] = a_0 + a_1 \cdot [A humidité g/kg]$
Humidité à partir de la matière sèche	(2)	$[A humidité g/kg] = a_0 + a_1 \cdot [A matière sèche g/kg]$

**Tableau 2.3**

Sommmation

**Table 2.3**

Sum

$$[A|C|U] = \sum a_i \cdot [A|C_i|U_i], \text{ avec } 0, i, n$$

Matière sèche délipidée (MSD)		$[A MSD U] = [A matière sèche U] - [A lipides U]$
Somme des acides gras saturés	(3)	$[A AG_{sat} U] = \text{somme } ([A AG_{sat(i)} U])$
Somme des AG mono-insaturés		$[A AG_{mono} U] = \text{somme } ([A AG_{mono(i)} U])$
Activité vitaminique A	(4)	$[A vitamine A g/kg] = [A rétinol g/kg] + [A carotène g/kg]/a_1$

**Tableau 2.4**

Transfert

**Table 2.4**

Imputation

$$[A|C|U] = a_0 \cdot [A_1|C|U] \cdot [A|C_1|U_1]^{a_1} \cdot [A|C_2|U_2]^{a_2}$$

Partie comestible (PC)	(5)	$[A Pc C U] = [A_{tel\ acheté} C U] \cdot a_0$
Rendement d'un traitement		$[A traité C U] = a_0 \cdot [A_{non\ traité} C U]$

(1) Les valeurs de  $k_p$ ,  $k_{AG}$  et  $k_b$  varient en fonction du type d'aliment.

(2)  $a_0 = 1\ 000$  et  $a_1 = -1$ .

(3)  $AG_{sat}$  et  $AG_{mono}$  sont respectivement les sommes d'acides gras saturés et mono-insaturés.  $AG_{sat(i)}$  et  $AG_{mono(i)}$  sont les acides gras dont le nombre de carbones varie entre 1 et 30.

(4)  $a_1$  vaut 6 sauf pour les produits laitiers pour lesquels il vaut 2.

(5)  $a_0$  représente la proportion de la partie comestible dans l'aliment tel qu'acheté ou l'effet d'un traitement, tel que la perte de vitamines pendant la cuisson.

C'est la solution adoptée par l'INFOODS, qui a proposé un *système de repérage* (tagname) des nutriments en vue de permettre un échange international des données (KLENSIN *et al.*, 1988). Pratiquement, c'est un code alphabétique comportant de deux à huit caractères et qui intègre l'unité de mesure (tableau 1).

Cependant, il existe un nombre important de nutriments qui sont obtenus en combinant différents constituants ou différents nutriments. Dans ce cas, un effort de normalisation reste à faire en ce qui concerne les méthodes de calcul et les valeurs des paramètres. Le meilleur exemple est celui de l'énergie obtenue par cumul des valeurs énergétiques des lipides totaux, protéines, glucides disponibles et alcool. Il existe de nombreuses variantes à ce calcul qui prennent plus ou moins en compte les glucides dits assimilables, les fibres, les acides organiques, etc.

Malgré tout, il apparaît que ces formules de calcul se regroupent en quatre grands types de modèles : le produit, le décalage, la sommation et le transfert. Si on prend comme convention de représenter chaque donnée consolidée par le triplet [A|C|U], où A symbolise l'aliment, C le nutriment et U l'unité de mesure, les tableaux 2.1 à 2.4 donnent quelques exemples de formules très classiques.

---

## 4 - LE SYSTÈME DE CODIFICATION DES ALIMENTS

---

### 4.1 Définition de l'aliment

Le contexte est plus délicat pour ce qui concerne les aliments car un nom d'aliment peut avoir plusieurs acceptions en fonction de références sociologiques, géographiques, etc. Une possibilité serait d'utiliser la définition légale, lorsqu'elle existe. Cette solution est intéressante mais elle présente quelques inconvénients. Par exemple, à propos de REGAL, il a été nécessaire de définir l'aliment dans un sens plus large que celui admis par la réglementation française. En effet, on considère comme un aliment toute substance pouvant être digérée et/ou assimilée par l'homme. Cette définition englobe bien sûr les matières premières, les produits transformés, les aliments composés et leurs ingrédients mais aussi les additifs. Il est indispensable d'inclure les additifs car ils ont une position particulière : selon le cas, un additif peut être regardé comme un aliment ou comme un nutriment. Par exemple, la molécule d'alpha-tocophérol déterminée par dosage dans un aliment est un nutriment ; par contre l'additif E307, ou alpha-tocophérol de synthèse, est un aliment puisqu'il peut intervenir comme ingrédient dans un aliment composé.

On estime, aujourd'hui pour un pays développé, à quelques milliers le nombre d'aliments différents qui auront à être gérés dans une banque de données sur la composition des aliments : ainsi les banques du Centre for Food Safety and Applied Nutrition (CFSAN) de la Food and Drug Administration (FDA) contiennent 15 000 noms d'aliments [Center for Food Safety & Applied Nutrition (CFSAN), U.S. Food and Drug Administration, 200 C Street S.W., Washington, D.C. 20204, USA].

**Tableau 3.1**  
Les facettes principales de LANGUAL

**Table 3.1**  
*Principal LANGUAL facettes*

Code	Facette	Définition
A	Type d'aliment	Famille ou groupe d'aliments défini à partir des propriétés de fabrication, d'utilisation ou encore des habitudes culturelles
B	Ingrédient principal	Espèce végétale ou animale ou produit chimique utilisé comme base de l'aliment
C	Partie utilisée	Sous-ensemble ou sous-produit de l'ingrédient principal utilisé pour fabriquer l'aliment (viande, lait, œuf, racine)
E	Etat physique	Etat se référant à l'aliment dans son entier (solide, liquide)
F	Traitement thermique	Sert à préciser si l'aliment a subi une cuisson quelconque lors de sa préparation (cru, cuit, ...)
G	Méthode de cuisson	Décrit la méthode utilisée (cuit dans l'eau, à la vapeur, par extrusion, ...)
H	Traitements technologiques	Tous traitements physiques ou chimiques employés pour modifier la composition de l'aliment Sert aussi à décrire les additifs, les auxiliaires de fabrication et les ingrédients
J	Préservation	S'applique au traitement primaire utilisé pour conserver l'aliment et décrit la technologie
K	Milieu de conditionnement	Précise le milieu physique employé lors du traitement de conditionnement (huile, sirop, azote)
P	Utilisateurs	Régime normal, hyposodé, etc.
Z	Particularités	Critère de qualification de la partie utilisée (label de qualité, noms des découpes de la viande)

**Tableau 3.2**  
Les facettes secondaires de LANGUAL

**Table 3.2**  
*Secondary LANGUAL facettes*

Code	Facette	Définition
M	Réceptacle ou emballage	Matériau et genre d'emballage utilisé
N	Surface en contact	Type de surface en contact avec l'aliment
Q	Lieu de prélèvement	Type d'établissement : fabricant, distributeur, détaillant, ...
R	Origine géographique	Lieu de production (pays, région)
S	Stockage	Durée et conditions de stockage
T	Période de production	Période ou année de production



Un numéro unique à cinq chiffres est donc amplement suffisant pour repérer chaque aliment. Attaché à ce numéro, on trouve une ou plusieurs dénominations dans une ou plusieurs langues. Ce numéro peut bien sûr être relié à d'autres nomenclatures (INSEE, Services Douaniers) à l'aide de tables de correspondance. Mais la création de ces tables de correspondance, surtout si elles sont internationales, pose des problèmes de cohérence : comment être sûr que les aliments repérés par chacun des systèmes à faire correspondre sont bien les mêmes ?

En effet, une simple dénomination ne permet pas une identification précise des aliments et risque de biaiser un échange d'informations. Les noms de certains aliments, tels que le gruyère ou le cheddar sont génériques et recouvrent une grande variété de produits selon le lieu de production. Des aliments d'origine locale ont des variantes dans chaque pays, afin de se conformer à la législation ou aux goûts locaux. Par exemple, on sait bien que si le mot pain peut se traduire par *bread* ou *Brot*, ce n'est pas pour autant que les pains anglais ou allemand sont fabriqués à partir des mêmes ingrédients qu'un pain français. Il apparaît ainsi que la traduction purement littérale d'un nom d'aliment est insuffisante pour appréhender de façon univoque sa composition chimique ou nutritionnelle.

Il faut donc avoir, à côté de la dénomination, une description claire, complète et scientifique. Un avantage complémentaire est qu'elle peut se prêter à des vérifications automatisées de similitude, ce qui simplifie les transferts de données entre banques.

Un tel système de codification descriptive des aliments est en cours d'élaboration entre plusieurs pays (FEINBERG *et al.*, 1988). Notamment, CODATA (Committee on Data for Science and Technology) a créé un groupe de travail sur ce sujet (Task Group on a Systematic Nomenclature for Foods in Numeric Databanks) qui fonctionne depuis 1989 (1).

## 4.2 Les principes du système LANGUAL

Le système de codification adopté s'appelle *LANGUAL* (Langua Alimentaria) et découle du système FFV (Factored Food Vocabulary) employé par la FDA (McCANN *et al.*, 1988 ; HEROLD, 1987 ; SMITH, 1989 ; PENNINGTON, 1989). *LANGUAL* est un langage normalisé de description, de classification et de recherche de données informatisées adapté aux aliments. Il est déjà utilisé dans plusieurs banques de données : aux Etats-Unis par la FDA (Total Diet Study et Scientific Information Retrieval and Exchange Network ou SIREN), par l'USDA (Nutrient Database for Standard Reference) et par le National Cancer Institute (Food Component Research Database) (BUTRUM et PENNINGTON, 1987) ; mais aussi au Canada par le Département Santé et Bien-être Social, au Danemark par la National Food Agency et en France pour la banque REGAL. Ce système est donc utilisable en anglais, en danois et en français.

(1) CODATA, comité interdisciplinaire scientifique du International Council of Scientific Unions (ICSU).  
CODATA secretariate : 51 boulevard de Montmorency, 75016 Paris.

Comme le système développé par l'International Network of Feed Information Center (INFIC), en usage mondial pour la description d'aliments pour bétail (HARRIS *et al.*, 1980 ; HARRIS, 1963 ; HAENDLER, 1988), LANGUAL est basé sur le principe d'un thésaurus en *facettes*. Chaque objet à coder est décrit par un ensemble de *termes standardisés*, eux-mêmes regroupés en *facettes*. Chaque *facette* représente un sous-ensemble de caractéristiques qui spécifient la qualité nutritionnelle et/ou hygiénique d'un aliment. Comme *facettes*, on rencontre l'origine biologique, les méthodes de conservation et de cuisson, les traitements technologiques, etc. Les tableaux 3.1 et 3.2 donnent la liste des 17 *facettes* utilisées jusqu'à présent dans le système LANGUAL. Dans le *tableau 3.1* figurent les *facettes* employées universellement alors que les *facettes* employées subsidiairement sont au *tableau 3.2*.

Il est évident que ces listes ne sont pas limitatives et que de nouvelles *facettes* peuvent être ajoutées, par exemple une *facette* spécifiant l'origine culturelle de l'aliment. C'est même l'intérêt d'un tel système, car l'ajout d'une *facette* ne remet pas en cause l'existant. Simplement, si cette *facette* doit être employée par plusieurs organisations, il est nécessaire qu'il y ait une concertation entre les différents utilisateurs pour préciser sa définition.

Chaque *facette* est repérée par un code alphabétique à une lettre. Le regroupement en *facettes* a été conçu pour simplifier la codification et la lecture des descriptions d'aliments en se basant sur les questions suivantes :

- A : à quel groupe d'aliments le produit appartient-il ?
- B-C : quelle est l'origine biologique de l'ingrédient principal ?
- E : quelles sont ses caractéristiques physiques ?
- F-J : quels ingrédients ont été ajoutés et quelles transformations ont permis d'obtenir le produit final ?
- K-N : quels sont son emballage et son milieu de conditionnement, quels matériaux sont en contact avec lui ?
- P : ce produit est-il destiné à des fins diététiques spécifiques ?
- Q-T : quels sont son origine géographique et son "âge" ?
- Z : a-t-il d'autres particularités ?

Grâce à ce système, un aliment analysé tel qu'acheté ou après préparation n'a pas les mêmes descripteurs (surtout pour les *facettes* E, F et G) et ne recevra pas le même numéro ; on pourra ensuite faire la différence entre les compositions de cet aliment cru et préparé. Le thésaurus peut être modifié en insérant ou en supprimant des descripteurs au sein d'une *facette*, ou même en rajoutant ou en éliminant une *facette* entière. Par exemple, les *facettes* R, S et T ont été créées par le CIQUAL et ne sont actuellement utilisées que dans une banque REGAL. De plus, lorsque des aliments nouveaux ou des technologies nouvelles apparaissent, il suffit de créer de nouveaux descripteurs pour prendre en compte ces aliments et ces procédés, sans pour autant bouleverser tout le système existant. C'est un système flexible et ouvert.

### 4.3 La structuration du thésaurus

Répartis entre ces 17 *facettes*, le thésaurus de LANGUAL contient plus de 2 500 descripteurs standardisés. Chaque descripteur possède une clé unique alphanumérique, formée de la lettre de *facette* et d'un numéro à quatre chiffres.

Ce numéro est choisi de façon purement aléatoire, au fur et à mesure que les descripteurs sont créés. On peut ainsi allonger la liste des descripteurs sans aucune difficulté, en fonction de l'évolution des aliments. La clé sert à repérer la dénomination du descripteur mais aussi sa définition. C'est, bien sûr, la mise au point de cette définition qui requiert le plus grand soin et le plus de travail, car elle doit être compréhensible et utilisable dans différents pays (voir § 4.4).

D'un point de vue documentaire, LANGUAL peut fonctionner comme un véritable système documentaire sans en avoir certains inconvénients. Il permet de construire facilement des équations de recherche basées sur des combinaisons de descripteurs utilisant l'algèbre de BOOLE mais il n'y a pas de contresens. Par exemple, si l'on veut retrouver tous les aliments contenant du lait, afin de vérifier leur teneur en lactose :

- avec un système documentaire classique on devra combiner un ensemble de mots-clés ;
- avec un système de codes numériques purement hiérarchique, il faudra créer manuellement une liste de numéros.

Dans les deux cas, on aura toujours des risques d'omissions. Par contre, grâce à ce système de description en facettes, on peut chercher les aliments dont le code descripteur du type de produit (facette A) est *produit laitier* (code A0164) et le croiser avec les codes C0113 (*lait ou extrait du lait*) et H0184 (*lait ajouté*) afin de ne pas omettre les aliments qui ne sont pas classés généralement dans les produits laitiers mais qui contiennent malgré tout des ingrédients à base de lait.

Les numéros des descripteurs sont, comme on l'a dit, attribués au hasard et chronologiquement. Cette méthode présente un inconvénient si l'on veut facilement raccorder deux descripteurs, par exemple pour élargir ou préciser une recherche. C'est pourquoi, une autre disposition a été intégrée à LANGUAL : les descripteurs d'une même facette forment une hiérarchie de termes-pères et de termes-fils. Ainsi, peut-on chercher les aliments décrits par le terme *gâteau* mais la recherche peut facilement s'étendre au père de ce descripteur (*produit de boulangerie ou de viennoiserie*) ou à son grand-père (*produit céréalier ou amylicé*) dans la mesure où leurs liens de parenté sont toujours connus.

Le tableau 4 donne un extrait de la hiérarchie des facettes B et H où les termes sont rangés par ordre hiérarchique ; les tabulations indiquent les niveaux. Dans cette liste on pourra ajouter, dès que ce sera nécessaire, de nouveaux descripteurs, comme *fermentation butyrique* ou *fermentation propionique* à la suite de *fermentation lactique*.

Cette hiérarchie entre les termes du thésaurus crée une certaine logique qui facilite la codification ou la recherche des aliments. Cependant, elle peut être trompeuse car elle est très liée à l'usage et à la culture. Ainsi, les anglo-saxons rassemblent sous le terme de *berries* plusieurs fruits qui apparaissent à des latins comme parfaitement séparés. D'un autre côté on parle en France de *fruits rouges* pour un sous-groupe qui n'a de sens que dans un contexte culturel français. Il semble donc préférable de créer des hiérarchies en fonction des habitudes nutritionnelles régionales sans vouloir développer une structure universelle.

**Tableau 4**  
Extrait de la hiérarchie de LANGUAL

**Table 4**  
The hierarchy of LANGUAL descriptors

---

**Facette B**

...		
B1139	Agrume	( <i>Citrus spp.</i> )
B1211	Cédrat	( <i>Citrus medica</i> )
B1751	Clémentine	( <i>Citrus clémentina</i> )
B1507	Pomélo	( <i>Citrus paradisi</i> )
B1028	pomélo et orange	
B1246	ananas Kumquat	
B1530	Kumquat	( <i>Fortunella spp.</i> )
B1275	Citron	( <i>Citrus limon</i> )
B1067	Lime	( <i>Citrus aurantifolia</i> ou <i>latifolia</i> )
B1339	Orange	( <i>Citrus sinensis</i> )
B1056	orange Valencia	( <i>Citrus sinensis</i> var. <i>Valencia</i> )
B1028	pomélo et orange	
B1709	orange et abricot	
B1631	ananas et orange	
B1193	orange Seville	( <i>Citrus aurantium</i> ou <i>bigaradia</i> )
B1898	orange bergamote	( <i>Citrus aurantium</i> ou <i>bigaradia</i> )
B2018	Pamplemousse ou shaddock	( <i>Citrus grandis</i> ou <i>decumana</i> )
B1473	Tangelo	( <i>Citrus tangelo</i> )
B1429	Mandarine ou tangerine	( <i>Citrus reticulata</i> )
B1502	Ugli ou pomélo-tangerine	

...

**Facette H**

...	
H0119	Modifié par des micro-organismes ou enzymes
H0128	Modifié par fermentation complexe
H0107	modifié par fermentation complexe essentiellement lactique
H0230	Fermenté par un procédé simple
H0256	fermenté au niveau des glucides
H0300	– fermentation acétique
H0232	– fermentation alcoolique
H0123	– fermentation alcoolo-acétique
H0101	– fermentation lactique
H0127	fermentation lipolytique
H0102	fermentation protéolytique
H0190	Mariné ou conservé dans le vinaigre
H0130	Modifié par un procédé physico-chimique
H0178	Aéré ou fouetté
H0206	Alcalinisé
H0197	Décoloré
H0276	Bromé
H0182	Glacé ou confit au sucre
H0270	Distillé
H9020	Extrudé
H0274	Mis en flocons
H0306	Homogénéisé ou émulsionné
H0174	Hydrogéné
H0277	Hydrolysé
H0121	Irradié
H0179	irradié par rayons ionisants
H0261	irradié par rayons UV
...	

---

#### 4.4 Les définitions associées aux descripteurs

Le succès d'un système de codification comme LANGUAL, en tant qu'instrument d'échange d'informations, dépend du degré de compréhension et du maintien de l'homogénéité des descripteurs. Dans ce but et pour se défendre des interprétations erronées, chaque descripteur est rattaché à une définition et une référence bibliographique. Le tableau 5 donne quelques exemples de ces définitions de descripteurs.

**Tableau 5**  
Exemples de définitions de descripteurs

**Table 5**  
*Examples of descriptor definitions*

---

A0106	<b>Produit préparé à partir de céréale ou de produit amylacé</b>
	<i>Définition</i> Aliment préparé par le séchage et/ou la cuisson d'un mélange de céréale broyée ou de produit amylacé, additionné d'un liquide et d'autres ingrédients tels qu'un corps gras ou édulcorant
	<i>Voir aussi</i> Produit de boulangerie, pâte alimentaire, céréales pour petit déjeuner
	<i>Référence</i> LANGUAL Scope-Notes 02 Oct. 1987
B1202	<b>Asiminier (<i>Asiminia triloba</i>)</b>
	<i>Définition</i> Plant Anonacée cultivée USA sud, Algérie, région niçoise Fruit mangé frais ou en dessert
	<i>Référence</i> CTIFL, "Nouvelles espèces fruitières", 1988
C0216	<b>Babeurre</b>
	<i>Définition</i> Liquide restant lorsque la crème est battue en beurre Peut contenir des petits globules de beurre
	<i>Voir aussi</i> Babeurre fermenté
	<i>Référence</i> LANGUAL Scope-Notes 02 Oct. 1987
H0173	<b>Salé</b>
	<i>Définition</i> A employer pour des produits renfermant plus de 4 % de chlorure de sodium naturel ou ajouté A employer aussi pour des produits salés, saumurés, conservés par le sel
	<i>Référence</i> LANGUAL Scope-Notes 02 Oct. 1987
H0177	<b>Noix ou graine ajoutée</b>
	<i>Définition</i> A employer pour un ingrédient à base de noix ou de graine (à l'exception des cacahuètes) qui se situe en deuxième position par ordre d'importance
	<i>Voir aussi</i> Cacahuète ajoutée
	<i>Référence</i> LANGUAL Scope-Notes 02 Oct. 1987

---

#### 4.5 La notion d'échantillon

Avec un système de codification en facettes, un problème à régler est de savoir à quelle limite raisonnable on peut fixer le nombre de facettes et la précision des descripteurs : doit-on décrire le *pain français* ou bien la *baguette fabriquée en 1989 par une boulangerie artisanale du nord de la France* ? L'une ou l'autre description ne représentera ni le même effort ni le même coût de codification. Pour tenir compte de cette limite de façon flexible on peut introduire, à côté de l'aliment, un autre concept, celui d'échantillon. Il représente, par exemple, le prélèvement effectivement analysé au laboratoire. Cette notion devient très utile si l'on veut gérer une collecte de mesures sur les contaminants.

Chaque échantillon est décrit à l'aide du même système de codification que l'aliment mais avec d'autres facettes. Dans REGAL, pour contrôler si un descripteur d'échantillon n'est pas employé à tort ou si on a utilisé assez de descripteurs, on a construit une table contenant le nombre de fois où chaque facette peut être employée.

**Tableau 6**  
La table de cardinalité

**Table 6**  
Cardinality table

Code	Facette	Mini	Maxi	Nombre
<b>Pour l'aliment</b>				
A	Type d'aliment	1	1	181
B	Ingrédient principal	1	1	1 170
C	Partie utilisée	1	1	157
E	Etat physique	1	1	45
F	Traitement thermique	1	1	5
G	Méthode de cuisson	1	2	33
H	Traitements technologiques	1	12	214
J	Préservation	1	2	46
K	Milieu de conditionnement	1	2	39
P	Utilisateurs	1	3	24
Z	Particularités	0	3	90
Sous-total				2 004
<b>Pour l'échantillon</b>				
M	Récepteur ou emballage	0	3	109
N	Surface en contact	0	3	43
Q	Lieu de prélèvement	0	1	5
R	Origine géographique	0	1	222
S	Stockage	0	2	3
T	Période de production	0	1	2
Sous-total				384
Total				2 388

On y trouve, pour chacune, un indicateur logique spécifiant si elle sert à décrire un aliment ou un échantillon, le minimum et le maximum de fois où elle peut être utilisée dans une codification. Cette table contenant le nombre d'éléments de chaque facette, c'est-à-dire le cardinal de l'ensemble au sens mathématique, est appelée *table de cardinalité*. Le tableau 6 résume son contenu et précise le nombre de descripteurs qui existaient au 1er janvier 1990. Grâce à cette table on sait que pour la facette H entre 1 et 12 descripteurs peuvent être utilisés pour un même aliment, parmi les 214 existants, et que la facette M est facultative pour un échantillon, son cardinal inférieur étant égal à 0. En outre, lors de la codification, il est toujours possible de contrôler si une description est complète ou non. Enfin une telle structure est parfaitement évolutive et permet de modifier facilement les règles de codification.

Un calcul rapide, à partir du tableau 6, montre qu'on peut utiliser entre 10 et 29 descripteurs pour un même aliment parmi les 2 004 déjà proposés. Le nombre de combinaisons possibles est donc énorme et l'expérience montre que plus de 15 000 aliments différents ont pu être décrits jusqu'à présent.

## 4.6 Règles de codification

Pour maintenir au mieux la cohérence de LANGUAL, un certain nombre de règles d'emploi des facettes ont été définies. Elles représentent l'état actuel d'achèvement de ce système, cependant elles nécessiteront une évolution et les tendances souhaitées sont signalées dans chaque cas.

### 4.6.1 Facette A

Le type d'aliment fait référence à ce que l'on appelle couramment la famille ou le groupe d'aliments. Les regroupements étant liés au fonction d'usage et au habitudes culturelles, le choix d'un descripteur A est toujours délicat. Il se fait en fonction des propriétés de fabrication (produit céréalier ou amylicé, fromage à pâte dure, ...) ou d'utilisation (boisson, ...) ou encore des habitudes culturelles (dessert, légume ou dérivé, ...).

Lorsqu'il y a doute, on code en fonction de l'utilisation. Par exemple, un jus d'orange est codé A0127 *jus de fruits*, mais une boisson à l'orange est codé A0112 *boisson non alcoolisée*. Les cacahuètes salées sont codées A0260 *amande ou noix ou graine ou dérivé*, mais les chips par A0228 *amuse-gueule*. Aujourd'hui dans REGAL, un seul descripteur A est autorisé pour décrire un aliment ; il sera sans doute nécessaire d'augmenter cette cardinalité.

### 4.6.2 Facettes B et C

L'origine de l'ingrédient principal (B) et la partie utilisée (C) indiquent la nature biologique de la matière première qui a servi à fabriquer l'aliment. C'est habituellement une espèce végétale (blé, champignon, carotte) ou animale (bœuf, poulet, thon) ou un produit chimique (eau, chlorure de sodium). Pour les denrées alimentaires composées de plusieurs ingrédients, on part de la liste pondérale décroissante des ingrédients (autre que l'eau) telle qu'elle apparaît sur la plupart des étiquettes pour définir l'ingrédient principal. Les autres ingrédients sont codés sous la facette H (traitements technologiques).

Par exemple, l'ingrédient principal du pain aux raisins est le blé (code B1421) et on code *fruit ajouté* sous la facette H. Il existe aussi quelques descripteurs B pour coder des mélanges les plus fréquents, comme *pomme et raisin* ou *porc et bœuf* mais leur usage est assez délicat.

La partie utilisée est souvent un produit de transformation ou un sous-produit de l'ingrédient principal qui a effectivement servi à fabriquer l'aliment, comme la feuille, le fruit, le muscle, le lait ou l'œuf. Dans l'exemple du pain aux raisins, il faut choisir le code C0208 *graine ou grain sans enveloppe et sans germe*, car c'est cette partie du blé qui sert à produire la farine blanche dont est faite ce pain. Dans l'exemple du jus d'orange, on code la plante B1339 *orange (Citrus sinensis)* et la partie utilisée C0167 *fruit*.

Si l'ingrédient principal est lui-même composite, on code son premier ingrédient. Par exemple, une barre de chocolat aux noisettes a comme ingrédient principal le chocolat au lait, et l'ingrédient principal de ce dernier est le sucre ; on choisit alors les codes B1012 *plante produisant du saccharose* et C0210 *saccharose*.

S'il n'y a pas d'ingrédient principal, comme dans le cas du gâteau Quatre-quarts (le même poids d'œufs, de farine, de sucre et de beurre), il faut coder B0001 *denrée alimentaire d'origine inconnue*, C0001 *partie de plante ou d'animal inconnue*, et ensuite indiquer les ingrédients sous la facette H (traitements technologiques). Dans le cas d'aliments complexes, cette règle de codification est insuffisante. Pour résoudre ce problème dans la banque REGAL, on a introduit le concept de *recette* qui permet de décrire sous la forme d'une liste d'ingrédients un aliment, tel un plat cuisiné, fabriqué à partir d'autres aliments. Une recette peut servir à décrire l'aliment, mais aussi à calculer des valeurs extrapolées des constituants. Cependant, des règles de gestion complémentaires doivent être construites pour éviter les références circulaires.

#### 4.6.3 Facette E

Il s'agit de l'état physique de l'aliment à 20°C, sans tenir compte de son éventuel milieu de conditionnement. Il peut être inhérent au produit ou résulter de traitements technologiques. Par exemple, à cette température, un sorbet est à l'état *liquide* et l'huile de coprah à l'état *solide*. L'état physique est important en hygiène alimentaire car il permet de signaler une prédisposition à des contaminations chimique ou microbiologique.

Pour cette facette, le principe de choix d'un descripteur est le suivant. D'abord on décide de l'état général du produit : *liquide peu visqueux* (eau, vin), *liquide très visqueux* (lait condensé), *semi-liquide* (miel), *semi-solide* (mayonnaise, fromage à tartiner), *solide* (pain). Ensuite, on peut choisir des descripteurs plus précis ; comme *solide tranché* (tranche de pain), *solide entier façonné* (miche de pain). Lorsqu'il y a doute, on choisit le terme le plus général. On n'admet qu'un seul descripteur de la facette E par aliment.

#### 4.6.4 Facettes F et G

Le traitement thermique sert à préciser si l'aliment a subi une cuisson quelconque lors de la préparation. Il faut préciser si la transformation thermique de l'aliment était *complète* (aliment cuit), *partielle* (aliment blanchi ou pasteurisé),



*inexistante* (aliment cru) ou inconnue. Si une partie de l'aliment est cuite et l'autre pas, on doit coder *partielle*. Un seul descripteur F est utilisé pour un aliment. Il faut signaler que l'interprétation de cette facette et le choix d'un descripteur sont très inspirés du droit alimentaire américain.

La méthode de cuisson décrit le processus d'application de la chaleur à l'aliment ou à ses ingrédients au cours de sa préparation qu'elle soit industrielle ou ménagère. Par exemple, pour coder un poulet rôti industriellement, puis refroidi et ensuite réchauffé dans un four à micro-ondes, on code G0005 *cuit ou rôti au four* et G0011 *cuit par micro-ondes*. Plusieurs descripteurs de la facette G peuvent être utilisés pour décrire un même aliment, dans le cas de REGAL on s'est limité à 2.

#### 4.6.5 Facette H

Les traitements technologiques décrivent les étapes de la préparation du produit qui impliquent l'addition, la substitution ou l'élimination de nutriments à l'aliment, ou bien les modifications technologiques de l'aliment ou de ses ingrédients. Plusieurs descripteurs de la facette H peuvent être utilisés pour un aliment.

Cette facette est utilisée pour indiquer les ingrédients secondaires d'un aliment composé, selon les règles indiquées dans la définition de chaque descripteur (voir *tableau 5*). Par exemple, le descripteur *aucun traitement appliqué* ne s'emploie que lorsqu'on a effectivement rien fait sur l'aliment ; le terme *traitement appliqué inconnu* est utilisé lorsque le traitement mis en œuvre n'est pas présent dans le thésaurus ou n'est pas connu. Par ailleurs, les traitements classiques qui n'auraient pas été employés, comme *sans adjonction de sel* pour du pain, sont indiqués sous le descripteur hiérarchique *traitement non appliqué*.

#### 4.6.6 Facette J

Le mode de préservation décrit le traitement primaire en usage pour prévenir la détérioration physique ou biologique du produit. Le descripteur J0131 *refroidi* ne doit être pris que dans le sens d'une méthode de conservation unique, comme pour le lait de grand mélange refroidi après la traite ; par contre, pour le lait pasteurisé réfrigéré, on choisit le code J0135 *pasteurisé par la chaleur*.

#### 4.6.7 Facette K

Le milieu de conditionnement est la matière dans laquelle l'aliment est éventuellement conditionné. En général, ce milieu peut être facilement séparé de l'aliment (fruits au sirop, sardines à l'huile). Les sirops sont classés selon la densité généralement signalée sur l'étiquette de l'aliment.

#### 4.6.8 Facettes M et N

Le récipient ou emballage fait référence, d'une part à la forme et d'autre part à la matière du récipient. La surface en contact spécifie les matériaux employés pour l'emballage qui arrivent en contact avec l'aliment. Plusieurs de ces descripteurs peuvent être utilisés pour décrire un même échantillon.

**Tableau 7.1**  
Pain français

**Table 7.1**  
French bread

Aliment		
A	0178	Pain
B	1421	Blé tendre ( <i>Triticum aestivum</i> )
C	0208	Grain sans enveloppe et sans germe
E	0105	Entier façonné épais de 0,5 à 7 cm
F	0003	Transformation complète par la chaleur
G	0005	Cuit au four
H	0107	Fermenté au niveau des glucides
J	0003	Sans méthode de conservation
K	0003	Sans milieu de conditionnement
P	0024	Produit de consommation courante
Echantillon		
M	0003	Sans récipient ou emballage
N	0003	Sans matière en contact avec l'aliment
Q	0002	Magasin ou détaillant
R	0250	France
S	0003	Sans stockage
T	0689	Production 06/89

**Tableau 7.2**  
Pain blanc américain

**Table 7.2**  
American white bread

Aliment		
A	0178	Pain
B	1418	Blé de force ( <i>Triticum aestivum</i> )
C	0208	Grain sans enveloppe et sans germe
E	0151	Solide
F	0003	Transformation complète par la chaleur
G	0003	Pas de méthode de cuisson applicable
H	0107	Fermenté au niveau des glucides
H	0136	Sucre ou sirop de sucre ajouté
H	0194	Enrichi ou à teneur garanti
H	0181	Enrichi en fer
H	0216	Enrichi en vitamine B
J	0001	Traitement de conservation inconnu
K	0003	Sans milieu de conditionnement
P	0024	Produit de consommation courante
Echantillon		
M	0001	Récipient ou emballage non spécifié
N	0001	Surface en contact avec l'aliment inconnu
Q	0001	Lieu de prélèvement inconnu
R	0840	USA
S	0001	Durée de stockage inconnue
T	1	Date de production inconnue

**Tableau 7.3**  
Camembert de Normandie au lait cru

**Table 7.3**  
Normandie camembert cheese from raw milk

Aliment		
A	0138	Fromage à pâte molle
B	1201	Vache
C	0245	Caillé du lait
E	0105	Entier façonné épais de 1,5 à 7 cm
F	0003	Sans transformation par la chaleur
G	0003	Pas de méthode de cuisson applicable
H	0107	Fermentation complexe essentiellement lactique
H	0298	Coagulant ajouté
H	0289	Affiné deux à quatre semaines
J	0003	Sans méthode de conservation
K	0003	Sans milieu de conditionnement
P	0024	Produit de consommation courante
Z	9001	Croûte fleurie
Z	0087	Appellation d'origine contrôlée (AOC)
Echantillon		
M	0177	Boîte en bois
M	0173	Emballage en papier
N	0039	Carton ou papier
Q	0004	Fabricant
R	0250	France
S	0003	Sans stockage
T	0488	Production 04/88

#### 4.6.9 Facette P

Les utilisateurs sont les groupes de consommateurs pour lesquels l'aliment est produit ou vendu : *aliment de premier âge* ou *aliment à teneur basse en sodium*. Le groupe de consommateurs est souvent mentionné sur l'étiquette de l'aliment. Plusieurs descripteurs de la facette P peuvent être utilisés pour décrire un aliment.

#### 4.6.10 Facettes Q, R, S et T

Le lieu de prélèvement (Q) indique le type d'établissement où l'échantillon a été pris (fabricant, distributeur, détaillant, ...). L'origine géographique (R) précise le lieu de production uniquement, sans tenir compte du lieu de consommation, et indique le pays et la région. La facette stockage (S) permet de relever la durée et les conditions dans lesquelles le produit a été entreposé avant prélèvement pour analyse (par exemple, *trois mois à 25°C*). La période de production (T) permet de noter la période ou l'année de production, pour décrire un aliment dont la composition est influencée, comme un lait de printemps ou un vin de 1989.

#### 4.6.11 Facette Z

Enfin, cette dernière facette regroupe les particularités complémentaires du produit, telles qu'un indicateur de qualité (AOC ou label, classes EUROP pour les viandes, maturité pour les fruits, ...), mais aussi les dénominations des découpes de viande, la finesse du hachage et le type de boyau des saucisses ou le type de croûte de fromage, ... Plusieurs descripteurs de la facette Z peuvent être utilisés pour décrire un même aliment.

---

### 5 - RÉSULTATS ET DISCUSSION

---

Trois exemples de codification d'aliments et d'échantillons sont présentés dans les tableaux 7.1 à 7.3. En particulier, on y compare les descriptions d'un pain produit en France et aux Etats-Unis. Le choix des descripteurs a été réalisé par les organismes américain et français qui développent LANGUAL, à savoir le CFSAN (FDA) et le CIQUAL. Les différences observées entre ces descriptions sont très intéressantes car elles peuvent expliquer une partie des désaccords entre tables de composition.

Plusieurs études ont été conduites pour comparer les compositions d'un "même" aliment présentées par les publications de plusieurs pays (POLACCHI, 1987). Elles concluent toutes à une grande variabilité due, en partie, aux erreurs systématiques introduites par les différentes méthodes d'analyse ou d'échantillonnage, mais aussi au fait que l'on a comparé ces aliments sur la base de leur appellation coutumière. Même des mesures de bonne qualité peuvent être en désaccord si elles proviennent d'aliments qui ne sont pas clairement identiques.

L'autre exemple porte sur le Camembert de Normandie au lait cru. Le choix de ce produit est justifié par le fait qu'il existe un label officiel sur ce type d'aliment. On voit ainsi comment LANGUAL permet de prendre en compte ces critères de qualité qui représentent un enjeu important dans le cadre de l'ouverture des marchés.

Comparé aux systèmes mono-hiérarchiques de codification des aliments, comme celui proposé par EUROCODE (ARAB *et al.*, 1987), le système de thésaurus à facettes semble être plus complexe mais en fait il est beaucoup plus flexible. En particulier il est bien adapté pour s'ouvrir à des denrées alimentaires nouvelles, à des traitements ou des caractéristiques qui n'ont pas été pris en compte jusqu'à présent. Il n'y a pas de limitations à la prise en compte de ces produits - ou catégories de produits - dès leur lancement sur le marché.

Néanmoins, le système LANGUAL ne peut pas être considéré comme accompli. Un problème majeur vient du fait qu'il a été conçu d'abord aux Etats-Unis, et beaucoup de descripteurs ont été développés pour leur relation étroite avec la législation et les habitudes alimentaires nord-américaines.

D'autres défauts proviennent de la classification des plantes et animaux qui ne suit pas la taxinomie. Ceci entraîne des limitations lorsque l'on veut coder les sous-parties d'un organisme. Par exemple, pour l'huile de foie de morue, on ne peut utiliser que les termes *huile* et *morue* sans pouvoir préciser davantage.

Des modifications et améliorations apparaîtront bientôt suite aux travaux du Groupe CODATA et de la plate-forme de coopération récemment mise en place au niveau européen dans le cadre du programme FLAIR (1) et enfin suite à l'extension de LANGUAL aux aliments africains. Malgré les efforts qu'ils exigent, ces systèmes de codification à facettes connaissent un succès certain dans le domaine agro-alimentaire, il suffit de se référer au code INFIC, utilisé mondialement, il constitue une clé commune pour tous ceux qui ont besoin de données sur la composition des aliments du bétail.

Le système LANGUAL permet donc de décrire, en détail et avec précision, les aliments actuels ou futurs, en provenance de divers continents ou cultures, de les codifier pour les retrouver aisément grâce à des systèmes automatisés de gestion. Associé à un système de codification des nutriments, harmonisé avec ceux des autres banques, LANGUAL donne ainsi la possibilité d'échanger sans ambiguïté des informations nutritionnelles avec les banques d'autres pays et de dessiner un éventuel réseau international de banques de données sur la composition des aliments.

En outre, le lien étroit qui existe entre LANGUAL et le droit alimentaire, à travers la définition des descripteurs, permet aussi de concevoir une harmonisation des réglementations qui ne sera pas seulement basée sur les normes de composition mais aussi sur des spécifications technologiques ou de moyens de production.

*Manuscrit reçu le 2 juillet 1990, accepté le 3 septembre 1990.*

---

(1) FLAIR, 1990. Improvement of the quality and compatibility of food consumption and composition data in Europe. Commission des Communautés Européennes, Bruxelles.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- ARAB L., WITTLER M., SCHETTLER G., 1987. *European food composition tables in translation*. Springer-Verlag, Berlin.
- BUTRUM R., PENNINGTON J., 1987. Technology systems used for food composition data bases In: *Computer handling and dissemination of data*, GLAESER P.S. (ed.), 404-408. Elsevier Science Pub., Amsterdam.
- FEINBERG M., FAVIER J.C., IRELAND-RIPERT J., 1987. *Répertoire général des aliments*. Tome 1: *Les corps gras*. Tome 2: *Les produits laitiers*. Lavoisier - Technique & Documentation, Paris.
- FEINBERG M., FAVIER J.C., IRELAND-RIPERT J., 1988. Problèmes liés à la codification des aliments: Le glossaire descriptif et analytique des aliments. *Congrès Aliment 2000, 23 janvier 1988, Paris*.
- FEINBERG M., IRELAND-RIPERT J., FAVIER J.C., 1991. Validated databanks on food composition: Concepts for modeling information. *World Review of Nutrition and Dietetics* (sous presse).
- HAENDLER H., 1988. Synthetic description systems for accurate data identification and selection: Principles and methods of nutritional data banks. *Int. Classif.*, 15, 2, 64-68.
- HARRIS L.E., 1963. A system for naming and describing feedstuffs, energy terminology, and the use of such information in calculating diets. *J. Animal Sci.*, 20, 535-547.
- HARRIS L.E., JAGER F., LECHE T.F., 1980. *International feed names and country feed names*. International Network of Feed Information Centers (INFIC), Logan, Utah.
- HEROLD P., 1987. *Using the factored food vocabulary in food composition databases*. National Cancer Institute (NCI-NIH) document, Bethesda, USA.
- HOLMAN P.C.H., KATAN M.B., 1985. *Report of the Eurofoods interlaboratory trial 1985 on laboratory procedures as a source of discrepancies between food tables*. State Institute for Quality Control of Agricultural Products, Wageningen.
- KLENSIN J.C., FESKANICH D., LIN V., TRUSWELL A.S., SOUTHGATE D.A.T., 1988. *Identification of food components for Infods data interchange*. IS N40. Massachusetts Institute of Technology, Cambridge, USA.
- LE GALLIC P., DUPUIS Y., BERNARDIN A., 1985. *Tables de composition des aliments*. Editions Jacques Lanore et Institut Scientifique d'Hygiène Alimentaire, Malakoff, Renaud.
- MCCANN A., PENNINGTON J.A.T., SMITH E.C., HOLDEN J.M., SOERGEL D., WILEY R.C., 1988. FDA's factored food vocabulary for food product description. *J. American Dietetic Assoc.*, 3, 336-341.
- PAUL A.A., SOUTHGATE D.A.T., 1978. *The composition of foods: McCance & Widdowson*. Elsevier, Londres.
- PENNINGTON J.A., 1989. Total diet study and factored food vocabulary: Languel. *14th National Nutrient Databank Conference, 19-21 Juin 1989*. University of Iowa, USA.
- POLACCHI W., 1987. Standardized food terminology: An essential element for preparing and using food consumption data on an international basis. *Food and Nutrition Bulletin*, 8, 2, 66-68.
- RANDOIN L., LE GALLIC P., DUPUIS Y., BERNARDIN A., 1960. *Tables de composition des aliments*. Lanore, Paris.
- RENAUD S., ATTIE M.C., 1986. *La composition des aliments*. INSERM et Astra-Calvé, Paris.
- SMITH E.C., 1989. Update on factored food vocabulary: Languel. *14th National Nutrient Databank Conference, 19-21 juin 1989*. University of Iowa, USA.
- USDA, 1976-1986. *Agriculture handbook No.8 series*. 16 tomes. Human Nutrition Information Service, Washington, USA.