

# Languag - harmonization of different national versions

## Report of the European Languag Technical Committee Meeting

Wageningen, the Netherlands, 20 October 1996

Report prepared by  
Florian Schlotke

Institute of Scientific Computing  
ETH Zentrum  
CH-8092 Zurich

Phone: +41-1-632 7458  
Telefax: +41-1-262 1172  
Email: schlotke@inf.ethz.ch

### Introduction

Languag has evolved over the last 20 years at US-FDA. More and more Languag is also used in Europe and other countries in the World. In order to become an international standard, existing versions of Languag need to be harmonized and a first official international version should be published by the International Languag steering Committee. This report contains the analysis, carried out by Florian Schlotke, of four existing versions of the Languag-thesaurus and suggests procedures to accomplish the publication of an official international thesaurus-version.

The Danish version has been ignored because it is outdated (in agreement with Dr. Anders Moller from the National Food Agency of Denmark). Due to organizational problems the French agency CREDOC didn't respond to the inquiry to send their version before this meeting in Wageningen.

The PC-software 'Thesaurus Manager' is an integral part of this report.

### Contents

1. Languag Version FDA 95.....	2
2. IIS 95 compared to FDA 95.....	4
3. CIQUAL 96 compared to FDA 95 .....	8
4. Hungary 96 compared to FDA 95.....	11
5. Summary and Conclusions.....	12
6. Future Work.....	13

## 1. Languag Version FDA 95

### 1.1 FDA 95

This version of the Languag-Thesaurus was sent from US-FDA to Anders Moller (National Food Agency of Denmark) in May 1995. It consisted of the two text-files 'vocabout.txt' containing all factors with their descriptors, synonyms and scope notes, and 'languag.txt' containing the tree-structure (or hierarchy) of the thesaurus.

A series of inconsistencies within these files were encountered. These were of the following three types:

1. Some factors of the B-facet referred to a broader-term that didn't exist in 'vocabout.txt' or 'languag.txt'. The following factors have been deleted, because they had only one, false reference to a broader term:

Factor	False pointer to broader-term
B2745 SLOE BERRY	B2745 ROSE FAMILY
B2746 ROWAN	B2746 ROSE FAMILY
B2747 UVA URSI	B2747 HEATH FAMILY
B2765 GRAPE IVY	B2765 GRAPE FAMILY
B2904 PARORE	B2904 SEA CHUB FAMILY

Other factors had more than one pointer to broader-terms. In this case, the false pointer was eliminated. These pointers were:

False pointer to broader-term
B2762 CHERIMOYA FAMILY
B2773 PEA FAMILY
B2802 MADDER FAMILY

2. Some factors in the file 'vocabout.txt' didn't have any pointer to a broader term. Fortunately the corresponding broader-terms could be found in the file 'languag.txt'. These references are now established:

Code	Corresponding Broader-Term
B1044	B1755
B1068	B1331
B1489	B2113
B1516	B1043
B1569	B1467
B1966	B1558
B1974	B1558
B2003	B1558
B2141	B1640
B2272	B1017
B2381	B1798
B2390	B1496
B2395	B1215

Code	Corresponding Broader-Term
B2422	B1581
B2438	B1295
B2440	B1625
B2446	B2032
B2463	B2330
B2543	B1365
B2614	B1081
B2643	B1640
B2645	B1526
B2647	B1856
B2658	B1581

3. There were two descriptors that appeared twice under different codes. The following redundant factors were deleted, because the codes '149' and '153' are not used anywhere else:

149 ANTIGUA AND BARBUDA

153 ARMENIS

These factors still exist under facet R with the codes 'AG' and 'AM' respectively.

Subsequently, this corrected version will be referred to as Languag-Thesaurus 'FDA 95'.

## 1.2 General observations

- The factors in the FDA Languag-Thesaurus often contain a French translation. These translations, however, are modeled as synonyms.
- The R-Facet: 'Geographic Places and Regions' breaks the 'rule' that each factor-code should consist of one letter indicating the facet followed by a four-digit number. The codes of facet R rather use ISO county-codes, sometimes followed by numbers indicating a region within a country. Groups of countries or continents use codes that start with the letter 'Q'. FDA also introduced four counties within north American states (e.g. US10A001 KENT COUNTY) forcing the Languag-code to be extended to at least 8 characters.
- The factor-codes contain leading zeros, that, depending on the facet, extend the code to a total length of three, four or five characters (e.g.: N04, G018, B0001). Since Languag uses arbitrary numbers to code factors we lose no information if we eliminate all leading zeros in the Languag codes (e.g.: N4, G18, B1). This was done in order to be able to compare FDA's thesaurus with the IIS-, French- and Hungarian-version.
- 7 of the 14 Languag-facets contain factors that refer to more than one broader-term. For example B1470 AVOCADO is listed under three categories: B1006 FRUIT USED AS VEGETABLE, B1024 TROPICAL OR SUBTROPICAL FRUIT and PIT FRUIT.

## 1.3 Facet overview

The following table gives an overview of the 14 facets of FDA 95.

FACET-FACTOR	FACET-TERM	Type	Factor s	Synonyms	References
A289	A. PRODUCT TYPE	S	190	315	189
B1564	B. FOOD SOURCE	M	1819	3979	1880
C116	C. PART OF PLANT OR ANIMAL	M	162	211	162
E113	E. PHYSICAL STATE, SHAPE OR FORM	S	46	87	45
F11	F. EXTENT OF HEAT TREATMENT	S	7	20	6
G2	G. COOKING METHOD	S	42	43	41
H111	H. TREATMENT APPLIED	M	269	282	277
J107	J. PRESERVATION METHOD	S	51	62	50
K20	K. PACKING MEDIUM	M	38	45	38
M100	M. CONTAINER OR WRAPPING	S	114	127	113
N10	N. FOOD CONTACT SURFACE	S	47	38	46
P32	P. CONSUMER GROUP/DIETARY USE/LABEL CLAIM	M	139	104	145
RQX	R. GEOGRAPHIC PLACES AND REGIONS	M	408	73	410
Z5	Z. ADJUNCT CHARACTERISTICS OF FOOD	M	214	121	216
<b>Total</b>			<b>3546</b>	<b>5507</b>	<b>3618</b>

The number of synonyms include the French synonyms. 'References' is the number of entries in the database tables that store the tree-structures of each facet. These figures give an impression of the size of each thesaurus. The 'Type'- column should be read as follows:

- S:** strictly hierarchic (or 'only one broader-term per factor allowed')
- M:** multiple broader-terms allowed

## 2. IIS 95 compared to FDA 95

### 2.1 IIS 95

IIS 95 is a Langual-version extracted from the August 1995 version of the 'International Interface Standard' software written by TAS, Inc. under contract for the FAD.

Because 31 duplicate factors and 11 factors that referred to themselves as broader terms were detected in the IIS 95 file, these were removed.

### 2.2 General observations

- A new facet has been added to Langual:

**Facet O: Cuisine**

This thesaurus is strictly hierarchic and contains 161 factors describing different ethnic 'cuisines'.

- Facet R. GEOGRAPHIC PLACES AND REGIONS has been tripled. Each descriptor now appears three times with the following prefixes: "PROCESSED IN ...", "ORIGIN IN ...", "CONSUMPTION OR PURCHASE IN ...". IIS 95 also put an R in front of each factor-code within facet R.
- It appears that IIS 95 evolved out of FDA 95. Some factors are missing, others inserted or moved. This also affects the tree-structure.
- IIS 95 does not provide any synonyms or scope-notes

### 2.3 Differences between IIS 95 and FDA 95

**No differences:**

There are no differences whatsoever within the following facets:

C. PART OF PLANT OR ANIMAL
E. PHYSICAL STATE, SHAPE OR FORM
F. EXTENT OF HEAT TREATMENT
G. COOKING METHOD
J. PRESERVATION METHOD
K. PACKING MEDIUM
M. CONTAINER OR WRAPPING
N. FOOD CONTACT SURFACE

**Minor differences:****Facet A: PRODUCT TYPE:**

The following factors appear in FDA95 and not in IIS95

A317 PRESERVATIVE  
A318 ANTIOXIDANT  
A319 STABILIZER, THICKENER OR GELLING AGENT  
A321 STABILIZER  
A322 ACIDIFIER

**Facet H: TREATMENT APPLIED:**

The following factors appear in FDA95 and not in IIS95

H394 ARTIFICIAL CHOCOLATE FLAVORING ADDED  
H395 ARTIFICIAL FRUIT FLAVORING ADDED  
H396 MARINADED  
H397 PLASMA ADDED  
H398 RIND REMOVED

**Facet P: CONSUMER GROUP/DIETARY USE/LABEL CLAIM**

The following Factor appears in FDA95 and not in IIS95

P182 ALIMENT DE DISETTE

**Facet Z: ADJUNCT CHARACTERISTICS OF FOOD**

Z205 SEMICULTIVTED PRODUCTION ENVIRONMENT

has been **added** as a new factor to IIS95 under Z152 AGRICULTURAL PRODUCTION ENVIRONMENT

Z155 TAIL

has two broader-terms in FDA 95:  
Z135 QUARTIER ARRIERE  
Z8 CUT OF MEAT, US  
The link to Z155 QUARTIER ARRIERE is missing in IIS95.

**Major differences:****Facet B. FOOD SOURCE**

There are 97 factors that appear in FDA95 and not in IIS95 and 7 factors that appear in IIS95 and not in FDA95. Moreover, some factors have been moved. In the following the differences in respect to FDA 95 will be discussed in related groups. NT is an abbreviation of narrower-term.

## B2395 ALGAE, BROWN

in FDA 95 a NT of B1215 ALGAE, BACTERIA OR FUNGUS USED AS FOOD SOURCE is on the same level as B1301 ALGAE. In IIS 95 it is a NT of B1301 ALGAE.

In this case FDA 95 is more recent than IIS 95, i.e. B1215 ALGAE, BACTERIA OR FUNGUS USED AS FOOD SOURCE is derived from B1215 ALGAE OR FUNGUS USED AS FOOD SOURCE, still used by IIS 95 and consequently lacks the following factors

## B2846 BACTERIA with all its Nts

- B2847 PSEUDOMONADALES USED AS FOOD SOURCE
- B2848 EUBACTERIALES USED AS FOOD SOURCE
- B2849 ACTINOMYCETALES USED AS FOOD SOURCE
- B2850 BEGGIATOALES USED AS FOOD SOURCE
- B2851 CYTOPHAGALES USED AS FOOD SOURCE
- B2852 SPIROCHAETALES USED AS FOOD SOURCE
- B2853 MYCOPLASMATALES USED AS FOOD SOURCE
- B2854 RICKETTSIALES USED AS FOOD SOURCE
- B2855 CHLAMYDOBACTERIALES AS FOOD SOURCE
- B2856 HYPHOMICROBIALES AS FOOD SOURCE
- B2927 CARYOPHANALES USED AS FOOD SOURCE

The following factors are also missing:

B2874 BROWN SHRIMP	B2877 GIANT STARGAZER
B2888 SIDESTRIPE SHRIMP	B2894 WHITE WAREHOUSE
B2880 COMMON PRAWN	B2866 EMPEROR FAMILY with its NT
B2925 PERIWINKLE FAMILY	B2859 EMPEROR
B1955 SEA SNAIL	B2886 CARDINALFISH FAMILY with all its NTs
B2881 PEN SHELL	B2885 BIGEYE CARDINALFISH
B2875 BLUE MUSSEL	B2936 CARDINALFISH
B2876 JAPANESE OYSTER	B2869 CUTTHROAT TROUT
B2879 RAZOR SHELL CLAM	B2860 ARCTIC GRAYLING
B2890 VENUS CLAM	B2870 POND SMELT
B2408 JELLYFISH	B2909 SLICKHEAD FAMILY with all its NTs
B2433 SEA CUCUMBER	B2907 LARGE SCALED BROWN SLICKHEAD
B2910 BAXTERS DOGFISH	B2908 SMALL SCALED BROWN SLICKHEAD
B2882 SHORTNOSE CHIMAERA FAMILY with its NT	B2914 BLACK SLICKHEAD
B2878 RATFISH	B2867 GRASS CARP
B2865 GREENLAND COD	B2862 SILVER DORY
B2864 BLUE LING	B2903 LOOKDOWN DORY
B2892 MORID COD	B2884 ALFONSINO FAMILY with its NT
B2899 GRENADIER FAMILY with all its NTs	B2883 ALFONSINO
B2897 GRENADIER	B2868 BUTTERFLY COD
B2898 DEEPSEA WHIPTAIL	B2893 ROCKFISH
B2905 KAIYOMARU RATTAIL	B2900 RED ROCKFISH
B2906 SMOOTH RATTAIL	B2916 FATHEAD FAMILY
B2912 HALFBEAK FAMILY with its NT	B2913 UNICORN RATTAIL
B2911 NEW ZEALAND GARFISH	B2696 SNIPE
B2918 SWOLLENHEAD CONGER EEL	B2917 SPELT
B2934 INDIAN MACKEREL	B2923 PSEUDOSPONDIA
B2935 NARROW-BARRED MACKEREL	B2926 AFRICAN LOCUST BEAN
B2891 SLENDER TUNA	B2928 DIALIUM
B2887 SCAMP	B2929 SALACIA
B2901 INCA SCAD	B2930 STRYCHNOS
B2896 YELLOWEYE MULLET	B2931 SWARTZIA
B2857 DENTEX	B2932 VANGUERIOPSIS
B2858 GILT HEADED BREEM	B2933 VITEX
B2871 PANDORA	B2061 GRAINS OF PARADISE
B2915 PALE TOADFISH	B2781 PITAYA
B2861 ATLANTIC CUTLASSFISH	B2873 LAND CRESS
B2902 SILVER SCABBARDFISH	B2889 BURDOCK, GOBO
B2872 BARRAMUNDI PERCH	B2920 BLACK PLUM
B2895 BLACK RUFF	

## B1244 LOQUAT

is not a NT of B1024 TROPICAL OR SUBTROPICAL FRUIT

### 3. CIQUAL 96 compared to FDA 95

#### 3.1 CIQUAL 96

This Languag-Version was provided by Jayne Ireland-Ripert of CIQUAL, France in Mai 1996.

#### 3.2 General observations

- CIQUAL 96 uses leading zeros so that all factor-codes, except in facet R, use one letter and a four-digit number.
- Facet R. GEOGRAPHIC PLACES AND REGIONS: Each factor-code starts with an 'R'. The codes are the same as in FDA 95. If a code is shorter than 5 characters, zeros are added at its end.
- 49 French factors have been **added**. They all have codes with numbers > 9000

Code	Descriptor	Broader-Term
A9018	rusk or industrial toast	A0178
A9901	puff pastry	A0202
A9902	choux pastry	A0202
A9903	shortcrust pastry	A0202
A9904	sablé pastry	A0202
A9905	strudel pastry	A0202
B9014	grape, chasselas	B2121
B9016	hare	B1134
B9017	black radish	B1315
B9026	gentian, yellow	B1179
B9027	pumpkin courge musquée	B1189
B9028	melonnette	B1190
B9028	potimarron	B1189
B9029	cornue d'hiver	B1190
B9030	patidou	B1190
B9031	patisson	B1190
B9032	pumpkin rouge vif d'Étampes	B1189
B9479	river trout	B1258
G9001	cooked in water or water-based liquid with pressure	G0013
G9002	vacuum cooked	G0021
H9013	pressed	H0130
H9249	no polyphosphates added	H0115
J9001	preserved by alcohol	J0145
P9008	label quality claim	P0059
RFR01	Ain	RFR00
RFR03	Allier	RFRB3
RFR15	Cantal	RFRB3
RFR21	Cote d'Or	RFRB4
RFR24	Dordogne	RFRB2
RFR33	Gironde	RFRB2
RFR40	Landes	RFRB2
RFR43	Haute Loire	RFRB3
RFR47	Lot et Garonne	RFRB2
RFR48	Lozere	RFRB3
RFR58	Nievre	RFRB4
RFR63	Puy de Dome	RFRB3
RFR64	Pyrenees Atlantiques	RFRB2
RFR67	Haut Rhin	RFRB1
RFR71	Saone et Loire	RFRB4
RFR75	Paris	RFRC7
RFR77	Val de Marne	RFRC7
RFR78	Yvellines	RFRC7
RFR89	Yonne	RFRB4
RFR91	Essonne	RFRC7
RFR92	Hauts de Seine	RFRC7
RFR93	Seine Saint Denis	RFRC7
RFR95	Val d'Oise	RFRC7
RFRV0	French wine growing regions	RFR00
RFRV1	Bordeaux	RFRV0

The following factors were **added** to FDA 95:

B1522 FRESHWATER CRAYFISH as a NT of B2686 CRAWFISH OR CRAYFISH  
B1053 CARRAGEENAN  
B1330 GHATTI  
B1363 TRAGACANTH  
B1445 STERCULIA as NTs of B1014 HEMICELLULOSE-PRODUCING PLANT

#### B2645 SILVER HAKE

has been **moved** from B1526 WEAKFISH to B1640 WHITING

#### B1204 DRUM:

NT B2863 MEAGRE is missing and B2430 TOTOABA has been **added** as a new NT

#### B2922 FICUS SPP.:

the following NTs:

B1302 FIG  
B2539 CLOWN FIG  
B2823 FICUS PLATYPHYLLA  
B2824 FICUS ITEOPHYLLA

are **rearranged** as

B1302 FIG  
B2539 CLOWN FIG  
B2823 FICUS PLATYPHYLLA  
B2824 FICUS ITEOPHYLLA

and **moved** to B1024 TROPICAL OR SUBTROPICAL FRUIT.

B2779 SYCAMORE FIG was **moved** to B1024 TROPICAL OR SUBTROPICAL FRUIT and finally B2922 FICUS SPP. is **missing**

#### B2921 EUGENIA SPP.:

is **missing** and its NT B2778 EUGENIA CAPENSIS **moved** to B1024 TROPICAL OR SUBTROPICAL FRUIT

#### B2468 KING ORANGE

was **copied** from B1429 MANDARIN ORANGE to B1339 ORANGE

#### B1753 GREEN BEAN AND PINTO BEAN

is not NT of B1371 GREEN BEAN

#### B1691 ROSELLE

was **moved** from B2082 HIBISCUS to B1024 TROPICAL OR SUBTROPICAL FRUIT



### 3.3 Differences between CIQUAL 96 and FDA 95

There are no factor-codes in FDA 95, that are not used in CIQUAL 96. But there are 44 factor-codes in CIQUAL 96 that do not appear in FDA 95 and that are not part of the French extension mentioned in chapter 3.2. According to Jayne Ireland-Ripert from CIQUAL, these factors have been added according to information from the US Languag technical Committee, and are more recent than the version FDA95. These factors are:

Code	Descriptor	Code	Descriptor
B1029	land snail family	H115	treatment not applied
B1053	carrageenan	H129	no preservative added
B1066	oriental tuna	H139	noncarbonated
B1073	escargot	H167	no color added
B1160	common sole	H170	unsweetened
B1330	ghatti	H176	unbleached
B1363	tragacanth	H193	no salt added
B1445	sterculia	H219	no flavor added
B1522	freshwater crayfish	H249	no nitrite/nitrate added
B1541	capsicum	H251	no water added
B1718	balsam pear	H255	not pasteurized
B1902	walking catfish	H258	unpurified, unrefined
B1937	butterfly kingfish	H264	less salt added
B1992	ethiopian lungfish	H269	no caffeine added
B2104	atlantic bluefin tuna	H312	no sulfite added
B2196	atlantic round herring	H313	no msg added
B2199	krill order	Z207	tail
B2240	louisiana crayfish		
B2371	medaka		
B2373	fish, cyprinodontiform		
B2430	totoaba		
B2477	stickleback		
B2494	perilla		
B2585	ocean sunfish		
B2619	ayu fish family		
B2626	krill		
B2765	grape ivy		

I was able to trace at least the following three factors and found some serious inconsistencies:

CIQUAL-Code	Descriptor	FDA-Code
B2626	krill	B2844
B2494	perilla	B2919
Z207	tail	Z155

These factors have the same descriptor but use different codes. This is due to entry-errors.

Synonyms are another potential source of differences or errors in the CIQUAL 96 thesaurus. This might be due to the relational schema used to store the thesaurus: it is not normalized, i.e. the synonyms and the tree-structure are not stored in separate tables.

The following is an example of a correct entry for a French synonym: all attributes but the French descriptor correspond.

LCode	Descriptor	French	Broader-Term
B2567	red mullet	rouget	B2658
B2567	red mullet	rouget barbet (Mullus barbatus ou surmuletus)	B2658

This is a example of an incorrect Languag-code:

A178 pizza crust is supposed to be A167

A178	bread	pain	A107
A178	pizza crust	pâte à pizza	A178

B1291 (hake) is not a BT of B1137 (pacific hake) in FDA 95. In FDA 95 'pacific hake' is just a synonym to 'Pacific Whiting'. Using one field for a descriptor (merlu du pacifique) and a synonym (Merluccius productus) might also lead to problems.

B1137	pacific hake	merlu du pacifique (Merluccius productus)	B1291
B1137	pacificwhiting	merlu du pacifique (Merluccius productus)	B1640

'sea cucumber' and 'sea slug' are synonyms of the same factor. Generally it is not a good idea to mix multiple synonyms and multiple broader-terms in the same two records. Moreover B2111 (gastropod) is not a broader-term of B2433 sea cucumber/sea slug in FDA 95.

B2433	sea cucumber	cornichon de mer (Holothuria spp.)	B2115
B2433	sea slug	beche de mer (Parastichopus californicus)	B2111

This is an examples of a correct implementation of multiple broader-term relations. It also adheres to FDA 95. But because the relational database-schema has not been normalised a **typing errors** occurred:

B1290	walnut	noix non spécifié (Juglans spp.)	B1017
B1290	walnut	noix non spécifiée (Juglans spp.)	B1062

Here is a similar case. The '?', however seems to be a remark: maybe somebody was not sure whether B1443 is really a BT of B2600

B2600	broccoflower	brocoli chou-fleur	B1094
B2600	broccoflower	brocoli chou-fleur ?	B1443

More typing errors:

B2068	field mustard	moutarde sauvage (Sinapsis ou Brassica arvensis ou sinapistrum)	B2069
B2068	field mustard	moutarde souvage (Sinapsis ou Brassica arvensis ou sinapistrum)	B2069

B2139	sago palm	sagoutier (Sagoutier metroxylon sagou ou Cucas cucinalis)	B1286
B2139	sago palm	sagoutier (Sagoutier metroxylon sagou ou Cucas cucinalis)	B1286

H280	sugar syrup or sugar syrup solids added	sirop de sucre ou solides de sirop ajouté	H136
H280	sugar syrup or sugar syrup solids added	sirop de sucre ou solides du sirop ajouté	H136

## 4. Hungary 96 compared to FDA 95

### 4.1 Hungary 96

Ivan Varsanyi of the University of Horticulture and Food Industry, Budapest provided a file of Hungarian foods coded with Languag. The Languag information in this file consisted of the factor-code and two versions of the factor-descriptor: the original FDA-version and a 'Hungarian'-version. The 'Hungarian'-version is also in English but sometimes it differs slightly to the original. I extracted all Languag-factors that were used to code the Hungarian foods and found 286 different factors. I didn't have any access neither to the other Languag-factors nor to the tree-structure used in Hungary.

### 4.2 Differences between Hungary 96 and FDA 95

All 286 factors used are available in FDA 95. 31 factors, however, differ slightly in their descriptor to the original FDA-version, because they have been copied from a former CIUQAL version:

Code	Hungary	FDA-Version of Descriptor
A138	Soft-ripened cheese (U.S.)	SOFT CHEESE (U.S.)
A150	Meat or meat product from mammal	MEAT OR MEAT PRODUCT (FROM MAMMAL)
A281	Semi-soft cheese (U.S.)	SEMISOFT CHEESE (U.S.)
B1043	Mackerel, kingmackerel	MACKEREL
B1055	Red and white currant	RED CURRANT
B1429	Tangerine	MANDARIN ORANGE
C139	Fruit or berry, peel present, core, core or seed removed	FRUIT OR BERRY, PEEL PRESENT, CORE, PIT OR SEED REMOVED
C140	Fruit or berry, peel present, core, core or seed present	FRUIT OR BERRY, PEEL PRESENT, CORE, PIT OR SEED PRESENT
C163	Fruit or berry, peel undetermined, core, core or seed present	FRUIT OR BERRY, PEEL UNDETERMINED, CORE, PIT OR SEED PRESENT
C189	Tongue (digestive system ?)	TONGUE
C213	Fruit or berry, peel undetermined, core, core or seed removed	FRUIT OR BERRY, PEEL UNDETERMINED, CORE, PIT OR SEED REMOVED
C229	Fruit or berry, peel removed, core, core or seed removed	FRUIT OR BERRY, PEEL REMOVED, CORE, PIT OR SEED REMOVED
C230	Fruit or berry, peel removed, core, core or seed present	FRUIT OR BERRY, PEEL REMOVED, CORE, PIT OR SEED PRESENT
C275	Whole animal, with skin, feathers or scales, not eviscera	WHOLE ANIMAL, WITH SKIN, FEATHERS OR SCALES, NOT EVISCERATED
F3	Not heat treated	NOT HEAT-TREATED
F14	Fully heat treated	FULLY HEAT-TREATED
F18	Partially heat treated	PARTIALLY HEAT-TREATED
F22	Heat treated	HEAT-TREATED
H107	Lactic acid-other organism fermented	LACTIC ACID-OTHER AGENT FERMENTED
H111	Treatment applied	H. TREATMENT APPLIED
H151	Spice added	SPICE OR HERB ADDED
J123	Sterilized by heat, canned	STERILIZED BY HEAT
J136	Frozen	PRESERVED BY FREEZING
N37	Polyvinyl chloride	POLY(VINYL CHLORIDE)
N39	Paperboard or paper	PAPER OR PAPERBOARD
P23	Human food, no age specification, special diet	FOOD FOR SPECIAL DIETARY USE
P24	Human food, no age specification, regular diet	HUMAN FOOD, NO AGE SPECIFICATION
P33	Human food, low calorie	LOW CALORIE FOOD
P37	Human food, reduced sodium	REDUCED SALT OR SODIUM FOOD
P39	Human food, low fat	LOW FAT FOOD
P48	Human food, dietary fiber claim	HIGH FIBER FOOD

## 5. Summary and Conclusions

FDA 95 is the only version that includes synonyms and scope-notes in an organized way. It also seems to be the predecessor of IIS 95. The French thesaurus evolved out of an earlier FDA-version and later updates of FDA 95. It is therefore recommended to take FDA 95 as a basis for the international 'Version 0' and all further discussions.

This 'Version 0' differs from FDA 95 in respect to the following modifications:

- correction of inconsistencies (see chapter 1.1)
- removal of French synonyms  
(a file of these synonyms can be obtained from the author on request)
- removal of leading zeros from the codes.

IIS 95 (e.g. the new facet O) and the French additions might serve as a pool of ideas and improvements to the Languag-thesaurus.

All versions contained errors because insufficient tools were used to manage the thesauri. All types of mistakes (dangling references, unreferenced factors, self-referencing factors, typing errors, duplicates, etc.) are not possible when a relational database-management system and a normalized database-schema is used.

Facet R seems to be a constant problem because of its code-structure. The idea to add a leading 'R' might help to better integrate this facet in the Languag canon. The issue of regions within a country (e.g. the US counties) also leads to confusion. The idea of IIS 95 to extend facet R to be applied as "PROCESSED IN ...", "ORIGIN IN ..." and "CONSUMPTION OR PURCHASE IN ..." is definitely useful in different applications, but we think this should be implemented in another way (IIS 95's suggestion leads to extra redundancy and to descriptors that are annoying for human readers).

Facet R is somehow unique: All facets (except R) state the context in which they may be applied in their title. Facet R just describes the **domain** (i.e. values that may be used) for different variables like PROCESSED\_IN or ORIGIN\_IN, whereas all other facets represent variables **and** domains, e.g. H. TREATMENT APPLIED.

## 6. Future Work

It is commonly agreed that a standardized international version of Languag should be published by the International Languag Steering Committee and will form the basis for all future discussions. This international version should be carefully designed, because no factor should be changed after publication. It will only be allowed to gradually extend the different facets (i.e. add or move factors). In order to develop Languag's first official international version, the European technical committee suggests the following procedure:

1. 'Languag version 0' (a cleaned version of FDA95) will be published, together with the 'Thesaurus Manager' software, on the World Wide Web (<http://food.ethz.ch:2000/languag/>). This will allow the different national agencies to translate the thesaurus. People who wish to download Languag (i.e. the thesaurus plus the 'Thesaurus Manager' software) are required to leave their name, address, etc. in order to keep track of where Languag is being used.
2. It was decided to hold a special meeting, sponsored by COST Action 99, February 1997 in Paris. The following topics should be discussed and long term decisions should be made:
  - Inclusion of INFOODS in the International Steering Committee
  - The format of Languag codes:
    - whether all codes should adhere to the same format (i.e. number of characters, leading zeros, facet indicated by first letter, ...)
    - procedures for examination and possible incorporation of national (temporary) factors into the international version.
  - Specification of rules, in special facet scope-notes, on how the thesauri are to be applied. These rules should be included in future coding and retrieval software. Such rules specify for example:
    - in which facets multiple broader-terms should be allowed (if at all);
    - whether root-descriptors for each facet are necessary;
    - in what context should descriptors be used (e.g. facet R).
  - The role and management of synonyms and scientific names. In Version 0, French synonyms have been deleted from FDA 95. It should also be discussed whether other types of relations commonly used in thesauri, e.g. 'related terms', should be supported in Languag.

