

Flexion des noms et des adjectifs composés : principes de codage

Michel Mathieu-Colas

► **To cite this version:**

Michel Mathieu-Colas. Flexion des noms et des adjectifs composés : principes de codage. 2011.
halshs-00635018

HAL Id: halshs-00635018

<https://halshs.archives-ouvertes.fr/halshs-00635018>

Submitted on 24 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FLEXION DES NOMS ET DES ADJECTIFS COMPOSÉS : PRINCIPES DE CODAGE

Avertissement. – Il ne sera pas fait de différence, ici, entre les diverses appellations des séquences figées : noms ou adjectifs *composés*, *locutions* nominales ou adjectivales, noms ou adjectifs *polylexicaux*, *lexies composées* et *lexies complexes* (Pottier), *synapsies* (Benveniste), etc. On conviendra de parler de « composés » dans un sens indifférencié pour désigner tous les syntagmes qui comportent, à quelque titre que ce soit, une part de figement.

Considérant que le TAL exige une reconnaissance précise de toutes les formes lexicales, nous avons élaboré une ressource linguistique aussi rigoureuse que possible et de large couverture. La description, naturellement, a commencé par les formes simples du français — soit, en comptabilisant toutes les formes fléchies, environ 500 000 graphies correspondant à plus de 760 000 valeurs. Le dictionnaire ainsi conçu, MORFETIK, est maintenant opérationnel et associe les données de base (sous forme de tables), un moteur de flexion, des interfaces de consultation et d’interrogation et un ensemble d’outils permettant la maintenance et l’exploitation des ressources¹.

Le même travail devait être entrepris pour les séquences figées (quel que soit le degré de figement). Aucun traitement informatisé n’est concevable si les unités complexes ne sont pas clairement identifiées, ce qui suppose un recensement très large et une description minutieuse. Nous avons donc élaboré, selon les mêmes principes que pour les mots simples, un dictionnaire des noms et des adjectifs « composés ». Ont été identifiés, d’ores et déjà, plus de 100 000 substantifs et plusieurs milliers d’adjectifs. La formalisation des tables est en voie d’achèvement².

Nous détaillons dans cette étude les particularités liées à la flexion de ces unités (mise au pluriel et, pour les adjectifs, au féminin), et plus précisément les principes de codage qui ont été adoptés pour pouvoir générer automatiquement, à partir de la liste des lemmes, toutes les formes fléchies correspondantes. La tâche, a priori, n’était pas évidente, compte tenu de la

1. Voir M. Mathieu-Colas, « *Morfetik* : une ressource lexicale pour le TAL », *Cahiers de lexicologie*, 2009-1, n° 94.

Cf. <http://halshs.archives-ouvertes.fr/docs/00/43/38/55/PDF/Morfetik.pdf>

Voir aussi P.-A. Buvet, E. Cartier *et al.*, « *Morfetik* : ressource lexicale pour le TAL », TALN’09, 2009 http://www-lipn.univ-paris13.fr/taln09/pdf/TALN_26.pdf

2. La description des composés, comme celle des mots simples, est le résultat de recherches menées par l’auteur depuis une vingtaine d’années au sein des laboratoires LLI et LDI. Le travail a consisté, d’une part, à recenser directement 75 000 unités complexes et à les compléter par d’autres sources, d’autre part, à élaborer, sur le plan méthodologique, le système de description morphologique.

diversité des formes (plus de 700 types morphologiques pour les seuls noms composés³). Il fallait compter aussi avec la complexité de certains schémas flexionnels, notamment pour les mots à trait d'union⁴ (FRANCS-*comtois* mais FRANC-*comtoises*, *arrière-GRAND-mères* ou *arrière-GRANDS-mères*). Nous n'avons pu y parvenir qu'en adoptant un système suffisamment flexible pour pouvoir prendre en charge tous les cas de figure. C'est ce système que nous présentons ici.

1. NOMS COMPOSÉS

Le codage morphologique des noms composés repose sur les mêmes principes fondamentaux que celui des noms simples, mais la complexité des formes implique un certain nombre d'ajustements.

La table des lemmes se présente ainsi :

nom	catgram	f1	f2	f3	f4	f5	f6	flexT	type	flexTot
abaisse-langue	nm							00;01	T	00;01
abaisse-paupière	nm							00	T	00
abandon de soi	nms								0	
abandon du domicile conjugal	nms								0	
à-bas	nm							00	T	00
abat-jour	nm							00	T	00
abat-son	nm							00;01	T	00;01
abat-sons	nm							00	T	00
abattement à la base	nm	01							I	+01/////
abat-vent	nm							00;01	T	00;01
abat-voix	nm							00	T	00
abbaye de monte-à-regret	nfs								0	
abbaye en commande	nf	01							I	+01/////
abcès sous-phrénique	nm	00		01					I	+00//01///

Rappelons que, pour des raisons à la fois théoriques et pratiques, les substantifs n'ont pas été fléchis en genre. On ne trouve donc dans la nomenclature que des noms masculins et des noms féminins.

3. Voir notre étude « Essai de typologie des noms composés français », *Cahiers de lexicologie*, 69, 1996-II, pp. 71-125. Cf. http://halshs.archives-ouvertes.fr/docs/00/43/38/41/PDF/Typologie_des_noms_composes.pdf

4. M. Mathieu-Colas, *Les mots à trait d'union. Problèmes de lexicographie informatique*, CNRS-INaLF, Didier érudition, coll. « Etudes de sémantique lexicale », Paris, 1994.

Ils sont codés ainsi, dans le champ « catgram » :

nm	nom masculin (flexion)	nf	nom féminin (flexion)
nms	nom masculin singulier	nfs	nom féminin singulier
nmp	nom masculin pluriel	nfp	nom féminin pluriel

à quoi s'ajoute le code **np** pour *bonnes / petites / vieilles / méchantes gens*, dont le fonctionnement en genre est très particulier.

Cela étant établi, il convient de décrire de manière précise le codage des flexions. Celles-ci peuvent relever de trois types : 0 (absence de flexion), T (flexion terminale) et I (flexion interne).

1.1. Type 0 (absence de flexion)

Certains noms peuvent être considérés comme figés en nombre (singulier ou pluriel obligatoire), d'où l'absence de code flexionnel :

<i>bien-être</i>	nms
<i>conscience de soi</i>	nfs
<i>beaux-parents</i>	nmp
<i>fines herbes</i>	nfp

1.2. Type T (flexions terminales)

Les mots de cette catégorie peuvent se rencontrer au singulier ou au pluriel, mais leur flexion fonctionne comme celle des mots simples :

- invariabilité :

un/des *laissez-passer*, une/des *pin up*

- ou flexion terminale :

un *sous-total*, des *sous-totaUX*
une *demi-heure*, des *demi-heureS*

Les codes sont indiqués dans le champ « flexT ». Le système de codage est le même que pour les mots simples, à l'exception de quelques formes qui ne se rencontrent qu'en composition :

28	-f > -ves	ice-shelF, ice-shelVES
29	-fe > -ves	jack-kniFE, jack-kniVES

(Voir en annexe 1 la table des codes de flexion nominale, complétée pour les besoins de la composition.)

En cas de variantes, les différents codes sont séparés par des points-virgules :

un *après-ski* des *après-ski* ou *après-skiS* 00;01
 un *après-match* des *après-matchS* ou *après-matchES* 01;25
 un *cross-country* des *cross-country*, *cross-countryS* ou *cross-countrIES* 00;01;27

nom	catgram	flexT	type
laissez-passer	nm	00	T
pin-up	nf	00	T
sous-total	nm	03	T
demi-heure	nf	01	T
ice-shelf	nm	28	T
jack-knife	nm	29	T
après-ski	nm	00;01	T
après-match	nm	01;25	T
cross-country	nm	00;01;27	T

1.3. Type I (flexions internes)

La description est ici plus complexe, dans la mesure où les noms correspondants peuvent avoir une ou plusieurs **flexions internes** :

assiette à soupe *assietteS à soupe*
artiste peintre *artisteS peintreS*
film interdit aux moins de dix-huit ans *filmS interditS aux moins de dix-huit ans*
conseiller prud'homal *conseillerS prud'homaUX*
passage à niveau non gardé *passageS à niveau non gardés*
arrière-grand-père *arrière-grandS-pèreS*

En conséquence, nous distribuons les codes sur plusieurs champs (f1 à f6) correspondant aux éléments constitutifs du composé :

nom	f1	f2	f3	f4	f5	f6	type
assiette à soupe	01						I
artiste peintre	01	01					I
film interdit aux moins de dix-huit ans	01	01					I
conseiller prud'homal	01		03				I
passage à niveau non gardé	01				01		I
arrière-grand-père		01	01				I

Les éléments sont identifiés séquentiellement, en tenant compte des séparateurs suivants : espace, trait d'union (ex. *arrière-grand-père*) et apostrophe (ex. *conseiller prud'homal*). Les codes sont les mêmes que pour les autres noms.

Gestion des variantes flexionnelles

Plusieurs cas peuvent se présenter :

a) Un même élément peut donner lieu à deux flexions distinctes :

match aller *matchS* ou *matchES aller*
double scotch *doubleS scotchS* ou *scotchES*

Comme pour les flexions terminales, ces variantes sont séparées par des points-virgules :

nom	f1	f2	type
match aller	01;25		I
double scotch	01	01;25	I

b) Un même élément peut être fléchi ou non fléchi :

ingénieur-conseil *ingénieurS- conseil(S)*
piano demi-queue *pianoS demi- queue(S)*
haricot vert sans fil *haricotS vertS sans fil(S)*

Les éléments facultatifs sont mis entre parenthèses :

nom	f1	f2	f3	f4	type
ingénieur-conseil	01	(01)			I
piano demi-queue	01		(01)		I
haricot vert sans fil	01	01		(01)	I

c) Quand il y a plusieurs parenthèses sur une même ligne, elles sont solidaires :

nom	f1	f2	f3	f4	type
chaussure à talon haut	01		(01)	(01)	I
périphérique entrée-sortie	01	(01)	(01)		I

Ce codage correspond aux pluriels :

chaussures à *talon haut* ou *talons hauts* (mais ni **talon hauts* ni **talons haut*)
périphériques *entrée-sortie* ou *entrées-sorties* (mais ni **entrée-sorties* ni **entrées-sortie*)

d) Lorsqu'une des variantes correspond à une flexion *terminale*, le codage est mixte. Voir le paragraphe suivant.

1.4. Type T+I (flexions internes et terminales)

Une même unité peut relever simultanément des deux types de flexion :

SINGULIER	FLEXIONS TERMINALES	FLEXIONS INTERNES
<i>fier-à-bras</i>	<i>fier-à-bras</i>	<i>fierS-à-bras</i>
<i>ampli-tuner</i>	<i>ampli-tunerS</i>	<i>ampliS-tunerS</i>
<i>social-démocrate</i>	<i>social-démocrateS</i>	<i>sociAUX-démocrateS</i>
<i>aller et retour</i>	<i>aller et retour</i>	<i>allerS et retourS</i>
<i>arrière-grand-mère</i>	<i>arrière-grand-mèreS</i>	<i>arrière-grandS-mèreS</i>

Nous associons dans ce cas les deux représentations :

nom	flexT	f1	f2	f3	type
fier-à-bras	00	01			T+I
ampli-tuner	01	01	01		T+I
social-démocrate	01	03	01		T+I
aller et retour	00	01		01	T+I
arrière-grand-mère	01		01	01	T+I

Ce système permet de prendre en charge certaines flexions complexes. En voici 3 exemples, avec les codages correspondants :

SINGULIER	FLEXIONS TERMINALES	FLEXIONS INTERNES
<i>fest-noz</i>	<i>fest-noz</i>	<i>festou-noz</i> <i>festoù-noz</i>
<i>reine-claude</i>	<i>reine-claude</i> <i>reine-claudeS</i>	<i>reineS-claude</i> <i>reineS-claudeS</i>
<i>surprise-party</i>	<i>surprise-partyS</i> <i>surprise-partIES</i>	<i>surpriseS-partyS</i> <i>surpriseS-partIES</i>

nom	flexT	f1	f2	type
fest-noz	00	A9;AA		T+I
reine-claude	00;01	01	(01)	T+I
surprise-party	01;27	01	01;27	T+I

1.5. Les composés *NAdj* et *N de N*

Le codage des noms de type *NAdj* (nom + adjectif) ou *N de N* (nom + de + nom), qui représentent à deux seuls plus des deux tiers des composés, s'inscrit aisément dans le format proposé.

1.5.1 Composés *NAdj*

Les singuliers et les pluriels obligatoires ne comportent pas de code de flexion :

mot	catgram	f1	f2	type
air comprimé	nms			0
acuité visuelle	nfs			0
abords immédiats	nmp			0
affaires courantes	nfp			0

Tous les autres relèvent du type « I » (flexion interne : f1 + f2), même en cas d'invariabilité de l'un des éléments (code « 00 » : voir *infra*, 1.7.1) :

mot	catgram	f1	f2	type
colis exprès	nm	00	00	I
abcès dentaire	nm	00	01	I
angle obtus	nm	01	00	I
angle aigu	nm	01	01	I
anneau nasal	nm	02	01;03	I
bail commercial	nm	04	03	I

A noter le cas exceptionnel de *droit commun* (« condamné de droit commun »), qui peut fléchir ou non les deux éléments : *des droit commun* ou (plus rarement en ce sens) *des droits communs*) :

mot	catgram	f1	f2	type
droit commun	nm	(01)	(01)	I

1.5.2 Composés *N de N*

Pas de code flexionnel pour les singuliers et les pluriels obligatoires (que le 2^e nom soit introduit ou non par un déterminant) :

nom	catgram	flex1	flex2	flex3	type
âge de bronze	nms				0
agilité d'esprit	nfs				0
arrêts de rigueur	nmp				0
absences de mémoire	nfp				0
air de la mer	nms				0

nom	catgram	flex1	flex2	flex3	type
abolition des privilèges	nfs				0
âges de la vie	nmp				0
allées du pouvoir	nfp				0

Les composés comportant un déterminant interne ne peuvent fléchir que le 1^{er} nom :

nom	catgram	flex1	flex2	flex3	type
as du volant	nm	00			I
accident de la circulation	nm	01			I
bécasseau des sables	nm	02			I
box des accusés	nm	00;25			I

Parmi les autres composés, certains ne fléchissent que le 1^{er} élément :

nom	catgram	flex1	flex2	flex3	type
abcès de fixation	nm	00			I
acte d'accusation	nm	01			I
match de barrage	nm	01;25			I

tandis que d'autres permettent une flexion du 2^e nom, mais de manière optionnelle (ce que nous notons, comme dans le cas général, par des parenthèses) : *des traitements de* TEXTE ou *des traitements de* TEXTES.

nom	catgram	flex1	flex2	flex3	type
cas d'espèce	nm	00		(01)	I
traitement de texte	nm	01		(01)	I
apport de capital	nm	01		(03)	I
barreau de chaise	nm	02		(01)	I

La flexion obligatoire des deux noms est tout à fait exceptionnelle :

nom	catgram	flex1	flex2	flex3	type
voix de châtré	nf	00		01	I

1.6. Codage synthétique

Le dernier champ de la table (« flexTOT ») reprend, de manière synthétique, l'ensemble des codages.

- type T : les flexions terminales sont reproduites telles quelles ;
- type I : les flexions internes sont précédées du signe +, les différentes zones de flexion (f1, f2... f6) étant séparées par des slashes ;
- type T+I : flexions terminales et flexions internes sont séparées par le signe « + ».

Nom	catgram	f1	f2	f3	f4	f5	f6	flexT	type	flexTOT
laissez-passer	nm							00	T	00
sous-total	nm							03	T	03
assiette à soupe	nf	01							I	+01/////
arrière-grand-père	nm		01	01					I	+01/01///
périphérique entrée-sortie	nm	01	(01)	(01)					I	+01/(01)/(01)///
aller et retour	nm	01		01				00	T+I	00+01//01///
reine-claude	nf	01	(01)					00;01	T+I	00;01+01/(01)////
surprise-party	nf	01	01;27					01;27	T+I	01;27+01/01;27////

1.7. REMARQUES

1.7.1. La différence entre le code « 00 » et l'absence de flexion est *purement théorique*. Pour des raisons linguistiques, nous codons :

<i>non-sens</i> :	flexion terminale	00	plur. <i>non-sens</i>
<i>faux sens</i> :	flexion interne	00-00	plur. <i>faux sens</i>
<i>sous-vêtement</i> :	flexion terminale	01	plur. <i>sous-vêtementS</i>
<i>bas-relief</i> :	flexion interne	00-01	plur. <i>bas-reliefS</i>
<i>timbre-poste</i> :	flexion interne	01-	plur. <i>timbreS-poste</i>
<i>grand bourgeois</i> :	flexion interne	01-00	plur. <i>grandS bourgeois</i>

Du point de vue informatique, les deux types de codage sont strictement équivalents.

1.7.2. Certains mots polysémiques se fléchissent différemment selon l'emploi. C'est ainsi que les composés de type « garde + Nom » n'ont pas, dans l'orthographe traditionnelle, le même comportement s'il s'agit d'humains ou de non-humains.

Il n'est pas tenu compte, ici, de cette distinction. Les flexions concurrentes sont provisoirement assimilées à des variantes, indépendamment de leur valeur sémantique.

2. ADJECTIFS COMPOSÉS

Le codage des adjectifs composés est encore plus complexe que celui des noms, dans la mesure où la **flexion en genre** vient s'ajouter à la mise au pluriel. En outre, certaines unités intègrent des déterminants ou des pronoms qui développent leur propre paradigme (*content de SOI, de MOI, de TOI, de LUI, etc.*).

Voici un extrait de la table des lemmes :

adj	catgram	f1	f2	f3	f4	f5	f6	flexT	type	flexTot
afro-asiatique	adj							40	T	40+/////
afro-brésilien	adj							50	T	50+/////
afro-cubain	adj							42	T	42+/////
agile comme un chat	adj	40		D4	40				I	+40//D4/40//
agile comme un écureuil	adj	40		D4	40				I	+40//D4/40//
agile comme un singe	adj	40		D4	40				I	+40//D4/40//
agité du bocal	adj	42							I	+42/////
agréable à l'oeil	adj	40							I	+40/////
agréable à l'oreille	adj	40							I	+40/////
agréable à vivre	adj	40							I	+40/////
agréable à voir	adj	40							I	+40/////
agréable au goût	adj	40							I	+40/////
agréable au toucher	adj	40							I	+40/////
agro-alimentaire	adj							40	T	40+/////
agro-industriel	adj							4D	T	4D+/////
agro-pastoral	adj							82	T	82+/////
aigre doux	adj	40	33						I	+40/33////
aigre-doux	adj	40	33						I	+40/33////
aigue-marine	adj							30	T	30+/////
aimable comme un bouledogue	adj	40		D4	40				I	+40//D4/40//
aimable comme un chardon	adj	40		D4	40				I	+40//D4/40//
aimable comme une porte de prison	adj	40		(D5)	(40)		(30;40)		I	+40//(D5)/(40)/(30;40)
aimé des dieux	adj	42							I	+42/////
air-air	adj							30	T	30+/////
air-sol	adj							30	T	30+/////
air-surface	adj							30	T	30+/////
al dente	adj							30	T	30+/////

La catégorie grammaticale est représentée par les codes suivants :

adj	adjectif normal (variable en genre et en nombre)
adjm	adjectif masculin (<i>avant-coureur</i>)
adjf	adjectif féminin (<i>belle comme une fleur</i>)
adjmp	adjectif masculin pluriel (<i>comme frère et sœur</i>)
adjfp	adjectif féminin pluriel (<i>sonnantes et trébuchantes</i>)

Pour ce qui est des flexions, on retrouve les 3 types précédemment identifiés : absence de flexion (0), flexion terminale (T) et flexion interne (I).

Les codes sont fondamentalement les mêmes que pour les adjectifs simples. En cas de flexion terminale (*infra*, 2.2), ils s'appliquent à la totalité du composé (*semi-public* se fléchit comme *public*). En cas de flexion interne (*infra*, 2.3), il s'appliquent aux éléments des composés (flexion de *royaliste* dans *plus ROYALISTE que le roi*).

Si certains codes propres aux mots simples ne se retrouvent pas ici (aucun adjectif composé ne fait apparaître l'alternance *s/ce* propre au code 32, que l'on trouvait dans *tiers/tierce*), quelques types de flexion, à l'inverse, ne se rencontrent qu'en composition :

- 41 *franc*-(maçon), *francs*-(maçons), *franc*-(maçonne), *franc*-(maçonnes)
- 83 *social*-(démocrate), *sociaux*-(démocrates), *social*-(démocrates), *social*-(démocrates) [variante flexionnelle]
- 96 (sans)-*égal*, (sans)-*égal*, (sans)-*égale*, (sans)-*égales* [variante flexionnelle]

On trouvera en annexe la table complète des codes.

2.1. Type 0 (absence de flexion)

Les pluriels obligatoires (catgram = adjmp ou adfp) ne donnent pas lieu à un code flexionnel. Cela ne concerne que quelques mots : *bons amis*, *comme frère et sœur*, *comme larrons en foire*, *en boules de loto*, *mari et femme*, *sonnantes et trébuchantes*.

2.2. Type T (flexions terminales)

De nombreux adjectifs composés se fléchissent comme des mots simples :

- invariabilité (féminin et pluriel identiques au masculin singulier) :

fair-play
à but lucratif

- ou flexion terminale :

oto-rhino-laryngologique, -es
à moitié vide, -es
semi-public, -cs, -que, -ques
sous-orbital, -aux, -ale, -ales

En cas de variantes, les différents codes sont séparés par des points-virgules.:

<i>anti-grève</i>	pl. <i>anti-grève</i> ou <i>anti-grèveS</i>	30;40
<i>semi-nasal</i>	pl. <i>semi-nasaLS</i> ou <i>semi-nasaUX</i>	42;82

adjectif	catgram	flexT	type
fair-play	adj	30	T
à but lucratif	adj	30	T
oto-rhino-laryngologique	adj	40	T
à moitié vide	adj	40	T
semi-public	adj	47	T
sous-orbital	adj	82	T
anti-grève	adj	30;40	T
semi-nasal	adj	42;82	T

2.3. Type I (flexions internes)

La flexion interne peut concerner n'importe quel élément :

- 1 ÂPRE *au gain*
- 1, 2 INCOLORE, INODORE *et sans saveur*
- 1, 2, 3 NOURRI, LOGÉ, BLANCHI
- 1, 3 NÉCESSAIRE *et* SUFFISANT
- 1, 4 NUL *et non* AVENU
- 2 *plus* ROYALISTE *que le roi*
- 2, 3 *mono-*OESTRIEN SAISONNIER
- 2, 4 *plus* MORT *que* VIF
- 3 *pas plus* HAUT *que trois pommes*
- etc.

Dans des cas semblables, nous distribuons les codes sur plusieurs champs (f1 à f6) correspondant aux éléments constitutifs du composé :

adjectif	f1	f2	f3	f4	type
âpre au gain	40				I
incolore, inodore et sans saveur	40	40			I
nourri, logé, blanchi	42	42	42		I
nécessaire et suffisant	40		42		I
nul et non avenu	4D			42	I
plus royaliste que le roi		40			I
mono-oestrien saisonnier		50	54		I
plus mort que vif		42		49	I
pas plus haut que trois pommes			42		I

Comme pour les noms composés, les éléments sont identifiés séquentiellement, en tenant compte des séparateurs suivants : espace, trait d'union, virgule et apostrophe.

2.4. Type T+I (flexions internes et terminales)

Une même unité peut cumuler les deux types de flexion (c'est le cas pour 56 mots) :

<i>double-crème</i>	pl. DOUBLE-CRÈME ou DOUBLES-CRÈMES
<i>en épingle à cheveux</i>	pl. en ÉPINGLE à <i>cheveux</i> ou en ÉPINGLES à <i>cheveux</i>
<i>grand ouvert</i>	fém. GRAND <i>ouverte</i> ou GRANDE <i>ouverte</i>
<i>papouan-néo-guinéen</i>	fém. PAPOUAN- <i>néo-guinéenne</i> ou PAPOUANE- <i>néo-guinéenne</i>
<i>tout de blanc vêtu</i>	fém. TOUT <i>de blanc vêtue</i> ou TOUTE <i>de blanc vêtue</i>

nom	flexT	f1	f2	f3	f4	type
double-crème	30	40	40			T+I
en épingle à cheveux	30		40			T+I
grand ouvert	42	42	42			T+I
papouan-néo-guinéen	50	42		50		T+I
tout de blanc vêtu	42	31			42	T+I

Exemple de flexion complexe : *social-démocrate*

Ce système permet de prendre en charge des variations complexes, comme dans le cas de *social-démocrate*. Selon les sources, *social-* reste invariable, ou ne varie qu'au masculin, ou varie normalement :

fém. sing.	masc. plur.	fém. plur.	codage
SOCIAL- <i>démocrate</i>	SOCIAL- <i>démocrates</i>	SOCIAL- <i>démocrates</i>	flexT = 40
SOCIAL- <i>démocrate</i>	SOCIAUX- <i>démocrates</i>	SOCIAL- <i>démocrates</i>	f1 = 83 ; f2 = 40
SOCIALE- <i>démocrate</i>	SOCIAUX- <i>démocrates</i>	SOCIALES- <i>démocrates</i>	f1 = 82 ; f2 = 40

ce qui peut se traduire ainsi :

adjectif	flexT	f1	f2	type
social-démocrate	40	82;83	40	T+I

Variation et polysémie

Certains adjectifs polysémiques se fléchissent différemment selon l'emploi. Par exemple, *garde-côte* n'a pas le même comportement selon qu'il se rapporte à des humains (grenadiers, canonniers, douaniers *garde-côtes* ou *gardes-côtes*) ou à des non-humains (bâtiments, vaisseaux *garde-côte* ou *garde-côtes*).

Il n'est pas tenu compte, ici, de cette distinction. Les flexions concurrentes sont assimilées à des variantes, indépendamment de leur valeur sémantique :

adjectif	flexT	f1	f2	type
garde-côte	30 ;40	40	40	T+I

2.5. Déterminants ou pronoms intégrés

Certains adjectifs intègrent des déterminants ou des pronoms, qui participent, au même titre que les autres éléments, au mécanisme de la flexion. Nous leur appliquons des codes spécifiques⁵.

2.5.1. Déterminants indéfinis (codes D1 à D5)

On comparera:

ms	fs	mp	fp	f1	f2	f3		f4
vêtu comme UN prince	vêtue comme UNE princesse	vêtus comme DES princes	vêtues comme DES princesses	42		D3		60
fait comme UN rat	faite comme UN rat	faits comme DES rats	faites comme DES rats	42		D4		40
myope comme UNE taupe	myope comme UNE taupe	myopes comme DES taupes	myopes comme DES taupes	40		D5		40

où D3 = *un, une, des, des* ; D4 = *un, un, des, des* ; D5 = *une, une, des, des*.

Ces formes apparaissent principalement dans les constructions en *comme*. Elles donnent souvent lieu à des variantes, selon que le comparant se trouve ou non fléchi :

sourd *comme un pot* plur. sourds *comme un pot* ou *comme des pots*

Ici encore, les flexions facultatives sont représentées par des parenthèses (les deux couples de parenthèses étant solidaires) :

adjectif	f1	f2	f3	f4	type
sourd comme un pot	42		(D4)	(40)	I

2.5.2. Déterminants définis (code D6)

Le code D6 (*le, la, les*) permet de fléchir un mot comme *le premier surpris* :

adjectif	f1	f2	f3	type
le premier surpris	D6	54	31	I

5. Voir aussi le cas de *un* dans « être fait l'un pour l'autre » : « Ils sont faits l'un pour l'autre » devient, au féminin, « elles sont faites l'une pour l'autre ». Le code 99 prend en charge cette flexion originale : *un* au masculin / *une* au féminin, au singulier (rare) comme au pluriel.

2.5.3. Déterminants possessifs coréférents au sujet (codes D7 à DC)

Certains possessifs n'apparaissent qu'à la 3^e personne :

ms	fs	mp	Fp	f1	f2	f3
réduit à SA plus simple expression	réduite à SA plus simple expression	réduits à LEUR plus simple expression	réduites à LEUR plus simple expression	42		D8

D'autres comportent en outre une flexion en personne :

ms	Fs	mp	Fp	f1	f2	f3
déchu de (MA, TA, SA, VOTRE) nationalité	déchue de (MA, TA, SA, VOTRE) Nationalité	déchus de (NOTRE, VOTRE, LEUR) nationalité	déchues de (NOTRE, VOTRE, LEUR) Nationalité	42		DB

2.5.4. Pronoms réfléchis (code PRO)

D'autres adjectifs intègrent le pronom réfléchi :

content de SOI, égal à SOI-même, fidèle à SOI-même, fier de SOI, hors de SOI, identique à SOI-même, imbu de SOI-même, livré à SOI-même, logique avec SOI-même, maître chez SOI, maître de SOI, plein de SOI-même, satisfait de SOI, satisfait de SOI-même, semblable à SOI-même, sûr de SOI.

Il en résulte une flexion spécifique, qui se traduit par un ensemble déterminé de formes que nous symbolisons par le code « PRO » :

ms	Fs	mp	fp	f1	f2	f3	f4
égal à (MOI, TOI, LUI, SOI, VOUS)-même	égale à (MOI, TOI, ELLE, VOUS)-Même	égaux à (NOUS, VOUS, EUX)-mêmes	égales à (NOUS, VOUS, ELLES)-mêmes	82		P2	40

Il est aisé de développer les 15 formes fléchies correspondant à cet exemple :

ms	Fs	mp	fp
je suis égal à moi-même	je suis égale à moi-même	nous sommes égaux à nous-mêmes	nous sommes égales à nous-mêmes
tu es égal à toi-même	tu es égale à toi-même	vous êtes égaux à vous-mêmes	vous êtes égales à vous-mêmes
vous êtes égal à vous-même	vous êtes égale à vous-même		
il est égal à lui-même	elle est égale à elle-même	ils sont égaux à eux-mêmes	elles sont égales à elles-mêmes
on est égal à soi-même			

2.6. Codage synthétique

Le champ « flexTOT » reprend, de manière synthétique, l'ensemble des codages, selon les mêmes principes que pour les noms (voir *supra*, 1.6) :

adj	catgram	f1	f2	f3	f4	f5	f6	flexT	type	flexTOT
fair-play	adj							30	T	30+/////
sous-orbital	adj							82	T	82+/////
âpre au gain	adj	40							I	+40/////
plus mort que vif	adj		42		49				I	+42//49//
frais émoulu	adj	30;39	42						I	+30;39/42////
déchu de sa nationalité	adj	42		DB					I	+42//DB///
égal à soi-même	adj	82		pro	40				I	+82//pro/40//
grand ouvert	adj	42	42					42	T+I	42+42/42////
tout de blanc vêtu	adj	31			42			42	T+I	42+31///42//
social-démocrate	adj	82;83	40					40	T+I	40+82;83/40////

ANNEXE 1 : CODES DE FLEXION NOMINALE

Chaque code correspond à un mode de génération du pluriel à partir de la forme du singulier. La colonne « S » indique les lettres du singulier qui doivent éventuellement être retranchées avant de rajouter les formes propres au pluriel (colonne « P »), par ex., pour le code 04 : *travaIL* > *trava-* > *travaUX*. Du point de vue informatique, il suffit d'indiquer le *nombre* de caractères à soustraire avant l'adjonction (colonne « Rad »). Le « radical » ainsi obtenu n'a pas de pertinence linguistique : il sert seulement de base à la génération automatique de la forme fléchie. A la limite, il peut s'agit d'une forme vide (comme dans le passage de *œil* à *yeux*, code 06). Les deux dernières colonnes illustrent chaque type de flexion, pour les noms masculins et/ou féminins.

Code	Rad	S	P	ex (nm)	ex (nf)
00	0			bras	souris
01	0		s	homme	femme
02	0		x	cheveu	peau
03	1	l	ux	amiral, ciel	
04	2	il	ux	travail	
05	2	il	ulx	ail	
06	4	oeil	yeux	œil	
07	1	l	aux	listel	
08	1	l	x	aïeul	
09	1	û	us	(port) dû	
0A	5	homme	shommes	bonhomme	
0B	5	adame	esdames		madame
0C	11	ademoiselle	esdemoiselles		mademoislle
0D	7	onsieur	essieurs	monsieur	
0E	10	onseigneur	esseigneurs	monseigneur	
0F	11	monseigneur	nosseigneurs	monseigneur	
10	0		a	lev	
11	1	i	a	méhari	
12	2	on	a	ostracon	
13	2	um	a	maximum	
14	0		e	leitmotiv, pfennig	nova
15	1	a	e		toccata
16	0		i	ricercar	
17	1	a	i		stotinka
18	1	e	i	ricercare	
19	1	o	i	pizzicato	
1A	1	u	i	leu	
1B	1	s	i	kouros	

Code	Rad	S	P	ex (nm)	ex (nf)
1C	1	s	ï	kouros	
1D	2	us	i	nævus	
1E	2	us	is	hippeus	
1F	0		en		Weltanschauung
20	2	an	en	gentleman	recorwoman
21	2	en	ina	foramen	
22	0		er	lied	
23	3	and	änder	land	
24	0		ler	il	
25	0		es	coach, conquistador	
26	1	i	es	guarani	
27	1	y	ies	lobby	lady
28	1	f	ves	ice-shelf	
29	2	fe	ves	jack-knife	
2A	2	ão	ões	sertão	
2B	1	á	ares		soleá
2C	2	ón	ones	colón	
2D	0		m	sefardi	
2E	0		im	kibboutz	
2F	1	e	im	ashkenaze	
A0	0		in	moudjahid	
A1	2	ed	adi	oued	
A2	2	ar	our	ksar	
A4	3	erm	oroum	cherm	
A5	2	ny	ce	penny	
A6	3	oot	eet	foot	
A7	2	al	é	bowal	
A8	2	al	és	bowal	
A9	0		ou	bagad	
AA	0		où	fest(-noz)	
AB	3	ets	tsy	tchervonets	
AC	3	etz	tsy	tchervonetz	
AD	5	argui	ouareg	targui	
AE	6	arguie	ouareg		targuie
AF	0		t		targuia

ANNEXE 2 : CODES DE FLEXION ADJECTIVALE

Les codes correspondent ici à la génération du féminin et du pluriel à partir de la forme du masculin singulier. Comme pour les noms, la colonne « Rad » d'indiquer le *nombre* de caractères à soustraire avant l'adjonction des formes propres au pluriel. La dernière colonne illustre chaque code par un exemple, étant entendu que la flexion porte spécifiquement sur l'élément en italique, sans préjuger du reste du composé (le code 34, par exemple, rend compte de la flexion de *amoureux* dans AMOUREUX *fou*, indépendamment de l'autre composant : AMOUREUX *fous*, AMOUREUSE *folle*, AMOUREUSES *folles*).

Code	Rad	MS	MP	FS	FP	exemples
00M	0			NULL	NULL	<i>gras</i> comme un chanoine
00F	0	NULL	NULL			<i>bas-bleu</i>
01M	0		s	NULL	NULL	<i>avant-coureur</i>
01F	0	NULL	NULL		s	<i>belle</i> comme une fleur
02M	0		x	NULL	NULL	<i>beau</i> comme un camion
03M	1	l	ux	NULL	NULL	<i>céphalo-spina</i> [l]
30	0					sans le sou
31	0			e	es	<i>épris</i> de justice
33	1	x	x	ce	ces	<i>dou</i> [x] comme un agneau
34	1	x	x	se	ses	<i>amoureux</i> [x] fou
35	0			se	ses	<i>bas</i> sur pattes
37	1	x	x	sse	sses	<i>fau</i> [x] comme un jeton
39	2	is	is	îche	îches	<i>fra</i> [is] et dispos
3C	2	ux	ux	ille	illes	<i>vie</i> [ux] comme le monde
40	0		s		s	<i>âpre</i> au gain
41	0		s			<i>franc-comtois</i>
42	0		s	e	es	<i>cousu</i> d'or
45	0		s	he	hes	<i>blanc</i> de peur
46	2	ec	ecs	èche	èches	<i>s</i> [ec] comme une trique
47	1	c	cs	que	ques	<i>semi-publi</i> [c]
48	0		s	que	ques	<i>judéo-grec</i>
49	1	f	fs	ve	ves	<i>vi</i> [f] comme la poudre
4B	0		s	ue	ues	<i>mi-long</i>
4D	0		s	le	les	<i>nul</i> et non avvenu
4E	1	u	us	lle	lles	<i>fo</i> [u] à lier
50	0		s	ne	nes	<i>bon</i> à rien
51	1	n	ns	gne	gnes	<i>mali</i> [n] comme un singe
52	3	ain	ains	ine	ines	<i>cop</i> [ain] comme cochon
54	2	er	ers	ère	ères	<i>avant-derni</i> [er]
55	2	et	ets	ète	êtes	<i>ultra-secr</i> [et]
59	0		s	te	tes	<i>sujet</i> à caution
60	0		s	sse	sses	<i>maître</i> de soi
6B	2	ur	urs	use	uses	<i>vole</i> [ur] comme une pie
6C	3	eur	eurs	rice	rices	<i>semi-conduct</i> [eur]
80	0		x		x	<i>peau-rouge</i>

Code	Rad	MS	MP	FS	FP	exemples
82	1	l	ux	le	les	médico- <i>léga</i> [l]
83	1	l	ux	l	l	<i>socia</i> [l]-révolutionnaire
85	2	au	aux	lle	lles	<i>be</i> [au] à voir
89	1	l	ux	l	ux	fort comme un <i>cheva</i> [l]
96	0			e	es	sans <i>égal</i>
98	0			e		<i>tout</i> en larmes
99	0			e	e	fait l' <i>un</i> pour l'autre
D1	2	un	des	NULL	NULL	beau comme [<i>un</i>] dieu
D2	3	NULL	NULL	une	des	belle comme [<i>une</i>] déesse
D3	2	un	des	une	des	tout seul comme [<i>un</i>] grand
D4	2	un	des	un	des	fait comme [<i>un</i>] rat
D5	3	une	des	une	des	myope comme [<i>une</i>] taupe
D6	1	e	es	a	es	[<i>e</i>] premier surpris
D7	3	son	leur	son	leur	à [<i>son</i>] comble
D8	2	sa	leur	sa	leur	réduit à [<i>sa</i>] plus simple expression
DA	3	mon ton son votre	notre votre leur	mon ton son votre	notre votre leur	avare de [<i>son</i>] temps
DB	2	ma ta sa votre	notre votre leur	ma ta sa votre	notre votre leur	déchu de [<i>sa</i>] nationalité
DC	3	mes tes ses vos	nos vos leurs	mes tes ses vos	nos vos leurs	perdu dans [<i>ses</i>] pensées
PRO	3	moi toi lui vous soi	nous vous eux	moi toi elle vous	nous vous elles	égal à [<i>soi</i>]-même