

Paul Sabatier

Laboratoire d'Informatique Fondamentale, Marseille
UMR 7279 – CNRS – Aix-Marseille Université

Paul.Sabatier@lif.univ-mrs.fr

Avril 2013

LVF+1

Une nouvelle version du dictionnaire électronique des verbes de Jean Dubois & Françoise Dubois-Charlier

Le travail présenté ici a été réalisé dans le cadre de FondamenTAL¹, une initiative pour le développement, l'enrichissement et la diffusion de ressources linguistiques fondamentales pour le traitement automatique des langues.

Note liminaire

Dans ce document, on fait référence aux fichiers (format Open Office) suivants :

- **LVF.ods** (Version originale du dictionnaire électronique des verbes)
- **LVF+1.ods** (Nouvelle version du dictionnaire électronique des verbes)
- **LVF+1_Predicats_Operateur.ods** (Explications sur la rubrique Opérateur)
- **LVF+1_Classes_semantiques.ods** (Explications sur les classes sémantiques)

qui sont téléchargeables sur le site où ce document a été téléchargé.

¹ <http://talep.lif.univ-mrs.fr/FondamenTAL>

DEV, LVF et LVF+1

On distingue :

- le « dictionnaire électronique des verbes (français) » (DEV) que Jean Dubois et Françoise Dubois-Charlier (D&D) ont réalisé en dBase et terminé en 1992 et qui a été diffusé à partir de septembre 2007 par MoDyCo² dans un format Excel ;
- l'ouvrage *Les verbes français* (LVF) que D&D ont publié en 1997 chez Larousse. Cet ouvrage étant épuisé, une reproduction au format PDF est diffusée par le laboratoire MoDyCo depuis septembre 2007.

L'ouvrage *LVF* est une présentation de certaines données contenues dans le DEV. Dans *LVF*, les verbes sont présentés par classes et sous-classes sémantiques. Dans le DEV, les verbes sont présentés par ordre alphabétique.

L'ouvrage *LVF* présente les différentes rubriques du DEV, à savoir : M(ot), DOM(aine), CLA(asse), OPER(ateur), SENS, PHRASE, C(onjugaison), CONST(ruction), DER(ivation), N(om), et L(exique).

Dans l'ouvrage *LVF*, pour chacune des entrées verbales on trouve la valeur de chacune des rubriques du DEV sauf les rubriques DOM(aine), C(onjugaison), N(om), L(exique) que l'on trouve seulement dans le DEV.

En ce qui concerne la rubrique DER(ivation), l'ouvrage *LVF* indique explicitement le mot dérivé (par exemple « lutte » pour « lutter 04 ») ou le ou les suffixes des noms et/ou adjectifs dérivés du verbe (par exemple « ure » pour « blesser 01 » donnant « blessure »). Dans le DEV, on trouve un ensemble de codes permettant de calculer le mot (nom et/ou adjectif) dérivé du verbe. Par exemple, pour « lutter 04 », on a dans le DEV le codage « --- -D ---- -- » et pour « blesser 01 », on a le codage « 11- -- ---- -A ».

Des algorithmes et des programmes ont été développés par Françoise Dubois-Charlier pour produire automatiquement à partir de ces codes les mots dérivés des verbes.

La distinction faite par D&D entre *LVF* (l'ouvrage *Les verbes français*) et le DEV (le dictionnaire électronique des verbes français) est donc pertinente. Cette distinction n'a pas toujours été faite jusqu'ici dans la littérature : certains auteurs parlent de *LVF* pour désigner aussi le contenu du DEV. Il semble aujourd'hui difficile de revenir en arrière. On peut appeler « *LVF* » le dictionnaire électronique des verbes en prenant soin de préciser que « *LVF* » désigne le dictionnaire électronique des verbes (DEV) dont l'ouvrage *Les verbes français* est une présentation d'une partie des données.

Dans ce qui suit, j'appellerai donc *LVF* le dictionnaire électronique des verbes français (DEV) dont l'ouvrage *Les verbes français* est une présentation d'une partie des données.

LVF+1 est une nouvelle version du dictionnaire électronique dans lequel j'ai apporté quelques correctifs, réécrit en clair des mots abrégés et certains codes, structuré la rubrique Opérateur et complété la rubrique Classe sémantique à partir des données contenues dans l'ouvrage *Les verbes français* de D&D.

² MoDyCo : <http://www.modyco.fr>

Rappelons que c'est pour des raisons d'espace mémoire que D&D ont dû recourir dès le début, il y a plus de 25 ans, à des chiffres et à des abréviations pour coder sur leurs ordinateurs et disquettes certaines informations dans leurs dictionnaires. Ils ont conservé ces codes bien établis dans leur systématique. Produire des versions de leurs dictionnaires avec des codes plus explicites et dans lesquels les abréviations sont remplacées par des formulations en clair est une tâche facilement réalisable, et cela pour deux raisons essentielles : d'une part parce que le codage est systématique, d'autre part parce que s'il y a ambiguïté pour interpréter au premier abord certaines abréviations, l'ambiguïté est levée en consultant d'autres rubriques pour l'entrée concernée.

Fichiers LVF.ods et LVF+1.ods

Le fichier **LVF.ods** contient la version originale du DEV dans laquelle j'ai ajouté une colonne intitulée <ID> contenant un numéro d'identifiant unique (un entier : de 1 à 25609) associé à chacune des 25 609 entrées du dictionnaire électronique. Le fichier **LVF+1.ods** est la nouvelle version que j'ai mise au point. On peut comparer les deux versions au moyen de la colonne <ID> commune aux fichiers LVF et LVF+1.

LVF+1

- **Des correctifs dans les rubriques**

J'ai apporté des correctifs qui ont été validés par Jean Dubois et Françoise Dubois-Charlier. Ces correctifs complètent ceux mentionnés dans le document daté du 11 janvier 2011³.

- **Rubrique MOT (M)**

Rappel : Cette rubrique contient les entrées du verbe à l'infinitif.

Dans LVF+1, les mots abrégés ont été réécrits en clair. Exemples :

évanouir 01 (s)	→	évanouir 01 (s')
promener 07 (s)	→	promener 07 (se)
fichu 01(ê)	→	fichu 01 (être)
languir 05(f.)	→	languir 05 (faire)
bienvenir (s f.)	→	bienvenir (se faire)
fouler 03 sol	→	fouler 03 le sol
férir (ss coup)	→	férir (sans coup)

- **Rubrique DOMAINE (DOM)**

Rappel : Cette rubrique précise au moyen d'un code pour les entrées verbales :

³ Ce document peut être téléchargé à l'adresse :
<http://pageperso.lif.univ-mrs.fr/~paul.sabatier/LVF/LVF-Correctifs.pdf>

- les domaines pragmatiques, techniques, scientifiques ou sphères d'application principales ;
- les niveaux de langue et les régionalismes.

Dans LVF+1, à partir de la rubrique Domaine, on a créé les trois rubriques : Domaine en clair, Nv Langue et Régionalisme qui formulent en clair le codage de la rubrique Domaine comme indiqué dans l'ouvrage *LVF* (p.12 et 17 de la version PDF de l'ouvrage)⁴. Exemples :

AER → aéronautique
 INS → instruments, appareils
 POL → politique
 VIT → viticulture, vin

f → familier p → populaire t → littéraire v → vieux

b → belgicisme s → helvétisme c → canadianisme, québécoïsme

- **Rubrique CLASSE (CLA) :**

Rappel : Cette rubrique code la classe sémantique à laquelle appartient l'entrée verbale concernée. Par exemple, 24 verbes sont classés C1d, c'est-à-dire comme appartenant :

- à la classe générique C des verbes de communication ;
- à la classe sémantico-syntaxique C1 : verbes de type « s'exprimer par un son, une parole » ;
- la sous-classe syntaxique C1d : verbes de type « parler, écrire à quelqu'un ».

C1d	24	"parler, écrire à qn"
-----	----	-----------------------

La classe C1d

L'ouvrage LVF spécifie en plus un sous-type pour 20 024 entrées (sur les 25 609). Par exemple, les 24 verbes de la classe C1d sont classés en 5 sous-types :

C1d.1	7	"parler à qn au moyen d'un appareil"
C1d.2	4	"parler à qn directement"
C1d.3	4	"parler à ou avec qn"
C1d.4	6	"s'adresser à qn"
C1d.5	3	"dire des paroles à qn"

Les cinq sous-types de la classe C1d

Il m'a semblé pertinent de préciser dans LVF+1 les sous-types comme cela est indiqué l'ouvrage LVF. Ainsi dans la rubrique Classe de LVF+1, le chiffre correspondant au numéro du sous-type (lorsque ce sous-type existe) est indiqué ; il est précédé d'un point, comme dans le tableau ci-dessus. Au final, la classification des verbes repose sur :

- 14 classes génériques ;

⁴ A été ajouté à la liste des domaines mentionnés à la page 12 de l'ouvrage *LVF* (version PDF), le domaine « spectacle » (SPE) présent dans le dictionnaire électronique.

- 54 classes sémantico-syntaxiques ;
- 248 sous-classes syntaxiques ;
- 533 sous-types.

Par ailleurs, on trouvera dans le fichier **LVF+1_Classes_semantiques.ods** qui accompagne ce document un tableau recensant l'ensemble des classes génériques, des classes sémantico-syntaxiques et des sous-types, avec leurs significations (formulées dans l'ouvrage *LVF*), des exemples de phrases (formulés dans l'ouvrage *LVF*), et le nombre de verbes concernés par la classification.

- **Rubrique OPERATEUR (OPER) :**

Rappel : Cette rubrique spécifie un schéma syntactico-sémantique qui définit l'emploi du verbe concerné, et cela au moyen d'un verbe de base et de ses compléments. Les schémas syntactico-sémantiques interprètent sémantiquement les schèmes syntaxiques formulés dans la rubrique Construction. Les verbes de base sont des abréviations de verbes basiques français, latins ou anglais. Ils permettent de classer les emplois verbaux dans des classes sémantiques structurées en sous-classes (Rubrique Classe).

Le contenu de la rubrique Opérateur obéit à une syntaxe précise. Dans *LVF*, Elle est formée par la suite de trois constituants (eux-mêmes structurables en sous-constituants) :

< sujet > < prédicat > < compléments et circonstants >

où < sujet > et < compléments et circonstants > peuvent être vides. Les constituants < prédicat > et < complément et circonstants > sont séparés par une espace. La dénomination de « prédicat » est personnelle.

Pour une meilleure lisibilité, j'ai noté dans *LVF+1* le constituant < prédicat > entre crochets. Exemple :

estiver 01 (an)f.mvt vrs pâture → (animal) [f.mvt] vers pâture

Le constituant < sujet >

< sujet > est de la forme (x) où x est un mot complet ou une abréviation. Chaque abréviation concernant le sujet a été réécrite en clair⁵, comme dans l'exemple précédent. Lorsque le sujet n'est pas précisé, nous l'avons noté (#). Il prend en général comme valeur celle du code du sujet spécifié dans la rubrique Construction.

Le constituant < compléments et circonstants >

Dans ce constituant tous les mots en abrégé ont été réécrits en clair⁶. La virgule (« , ») a été remplacée par la barre oblique (« / ») qui se lit « ou ». Exemples :

⁵ Sauf les abréviations suivantes : *qc* (quelque chose), *qc* (quelqu'un), *som* (corps, partie du corps), *log* (forme/production langagière écrite ou orale), *+pl* (+pluriel).

⁶ Sauf les abréviations suivantes : *qc* (quelque chose), *qc* (quelqu'un), *som* (corps, partie du corps), *neg* (négation), *+pl* (+pluriel).

abhorrer	sent haine pr qn,qc	→ (#) [sent] haine pour qn/qc
absorber 03	(qc)grp bruit,lux	→ (qc) [grp] bruit/lumière
adresser 02	f.ire qn A qn,instit	→ (#) [f.ire] qn A qn/institution

Dans le constituant <compléments et circonstants> on distingue les prépositions écrites en majuscules et celles écrites en minuscules. Cette distinction a toute son importance : les prépositions en majuscules introduisent les compléments prépositionnels du prédicat, préposition que l'on peut retrouver dans le codage des schèmes syntaxiques formulés dans la rubrique Construction associée à l'entrée verbale concernée. Nous avons réécrit en clair les prépositions en majuscules, comme indiqué dans le tableau ci-dessous.

LVF	LVF+1
AV	AVEC
AV/P	AVEC/PAR
CT	CONTRE
CT/PR	CONTRE/POUR
CT/SR	CONTRE/SUR
D	DE
D/AV	DE/AVEC
D/SR	DE/SUR
D-AV	DE/D'AVEC
DS	DANS
E	EN
P	PAR
PR/CT	POUR/CONTRE
SR	SUR
SR/CT	SUR/CONTRE
VRS	VERS

Prépositions liées aux prédicats (Rubrique Opérateur)

La barre oblique (« / ») se lit « ou ». Plutôt que de réécrire D-AV en DE-AVEC, nous avons préféré réécrire DE/D'AVEC qui est plus explicite. A distinguer bien sûr de la réécriture de D/AV en DE/AVEC. J'ai volontairement maintenu la distinction entre CT/PR et PR/CT. Le premier (8 entrées concernées) porte sur « li.mut » ; le second (21 entrées concernées) porte sur « loq ». De même, j'ai maintenu la distinction entre CT/SR et SR/CT. Le premier (33 entrées concernées) porte sur « f.mvt », « grp » et « loq » ; le second (64 entrées concernées) porte sur « f.ire », « ict » et « lc ».

Le constituant Prédicat

Comme indiqué plus haut, pour une meilleure lisibilité, nous avons noté le prédicat entre crochets.

Le **prédicat** se compose :

- d'une tête, **la tête du prédicat**, constituée d'un seul élément ou de plusieurs éléments alors séparés par un point ;
- d'une éventuelle suite de **modifieurs** préfixés par le signe « + ».

Exemples de têtes de prédicats :

abda	ag	av.car	dat	dat.val dli	dli.accord	dli.clo
m.e.accord		m.e.état	m.e.marche	percep	percep.mens	

La question s'est posée de savoir si les têtes des prédicats devaient ou non être réécrites en clair. Après une étude approfondie du dictionnaire, j'ai préféré conserver les intitulés des têtes de prédicats. En effet, si dans certains cas on pourrait réécrire en clair la tête de prédicat, comme par exemple :

m.e.mvt → mettre.en.mouvement

il s'avère que dans d'autres très nombreux cas, la réécriture d'une même tête de prédicat pourrait varier selon l'entrée verbale concernée. Opérer une telle réécriture conduirait à perdre le sens générique exprimé dans le verbe de base.

Pour une compréhension plus immédiate des prédicats de la rubrique Opérateur, on consultera le fichier **LVF+1_Predicats_Operateur.ods** qui accompagne ce document. Il s'agit d'un tableau composé des rubriques suivantes :

- LVF (Prédicat OP) : Liste des prédicats dans la rubrique Opérateur de LVF original ;
- LVF+1 (Prédicat OP) : Liste des prédicats dans la rubrique Opérateur de LVF+1 ;
- Simpl./Comp. : Prédicat simple (S) / Prédicat composé (C) ;
- LVF:Nb : Nombre d'occurrences du prédicat dans LVF original ;
- LVF+1:Nb : Nombre d'occurrences du prédicat dans LVF+1 (après correctifs) ;
- Classe : Classe sémantique où apparaît le prédicat ;
- Signification : Signification du prédicat.

Dans le constituant <prédicat>, des *modifieurs* peuvent porter sur la tête du prédicat. Les modifieurs sont les suivants :

LVF	LVF+1	Signification
+attribut	+attribut	attribut
+durée	+durée	durée
+fut	+futur	futur
+ind	+indicatif	indicatif
+ind/D	+indicatif/DE	indicatif ou infinitive en DE
+ind/inf	+indicatif/infinitif	indicatif ou infinitif
+inf	+infinitif	infinitif
+inf,qc	+infinitif/qc	infinitif ou quelque chose
+nég	+nég	négation
+pp	+pp	participe passé
+ql	+ql	qualitatif
+qt	+qt	quantitatif
+re	+re	répétition
+subj	+subjonctif	subjonctif
+subj/D	+subjonctif/DE	subjonctif ou infinitive en DE
+tps	+tps	temps

Modifieurs (Rubrique Opérateur)

Certains de ces modificateurs peuvent apparaître aussi dans les composants < sujet > et < compléments et circonstants > de la rubrique Opérateur.

Rubrique SENS

Rappel : Cette rubrique présente des parasyonymes ou des formes abrégées de définitions qui éclairent éventuellement le sens de l'entrée verbale concernée.

Dans LVF+1, les mots abrégés ont été réécrits en clair⁷. Exemples :

abriter 01	se m à l'abri	→ se mettre à l'abri
démontrer 04	f démonstration d'appar	→ faire démonstration d'appareil
entendre 01	jouir de b/mvs audition	→ jouir de bonne/mauvaise audition
iriser 01	donner color arc-e-ciel	→ donner couleur arc-en-ciel
soutirer 02	obtenir arg de qn	→ obtenir argent de quelqu'un

Dans cette rubrique figurent, dans la version originale, pour certains verbes des variantes orthographiques notées entre parenthèses, variantes qui font l'objet d'entrées dans le dictionnaire. Les variantes en abrégées ont été réécrites en clair. Toutes les variantes ont été placées entre crochets. Exemples :

débudgéter	ôter du budget(-tiser)	→ ôter du budget [débudgétiser]
débudgétiser	ôter du budget(-géter)	→ ôter du budget [débudgéter]
sabler	couvrir de sable(-onn-)	→ couvrir de sable [sablonner]
sablonner	couvrir de sable(-bler)	→ couvrir de sable [sabler]

Les parenthèses ont été conservées pour les compléments d'informations originaux ou que nous avons précisés. Exemples :

claver	serrer avec clavus(clé)	→ serrer avec clavus (clé)
saponifier	changer en savon(sapo)	→ changer en savon (sapo)
chondrifier (se)	se m en chondr-cartila	→ se mettre en chondr (cartilage)
zader	donner statut de zad	→ donner statut de ZAD (zone d'aménagement différé)

- **Rubrique PHRASE**

Rappel : Cette rubrique présente une ou plusieurs phrases simples illustrant l'emploi de l'entrée verbale concernée.

Dans LVF+1, les mots abrégés ont été réécrits en clair⁸. On a conservé la lettre P comme abréviation d'un prénom (ou d'un nom propre) humain. Exemple :

abrutir 02 On a~ P de questions.

On a appliqué les règles typographiques en matière de ponctuation. Exemple :

⁷ Sauf les abréviations suivantes : *qn* (quelqu'un), *qc* (quelque chose), *qp* (quelque part), *som* (corps, partie du corps).

⁸ Sauf l'initiale (de l'infinif : a~, b~, ..., z~) du verbe concerné dans l'exemple.

abaisser 04

On a~les prix,les revenus de dix p.c.Les prix s'a~de bcp. →

On a~ les prix, les revenus de dix pour cent. Les prix s'a~ de beaucoup.

- **Rubrique CONJUGAISON (C) :**

Rappel : Cette rubrique code le modèle de conjugaison et l'auxiliaire pour l'entrée verbale concernée. Les codes sont précisés page 23 du livre *LVF* (version PDF).

Aucun correctif n'a été apporté pour l'instant dans cette rubrique.

- **Rubrique CONSTRUCTION (CONST)**

Rappel : Cette rubrique code les schèmes de construction syntaxique de l'entrée verbale concernée. Les codes sont précisés pages 11-12 du livre *LVF* (version PDF).

Pour les entrées verbales *découpler 01*, *déparier 02*, *désaccoupler* et *désapparier 01*, le correctif suivant a été apporté : T1700 → T1200.

Lorsqu'une entrée verbale présente plusieurs constructions, nous les avons séparées par un tiret. Exemples :

A11 T1300	→	A11 - T1300
N1a A10 P7000	→	N1a - A10 - P7000
N1d A70 P7000 T1100	→	N1d - A70 - P7000 - T1100

- **Rubrique DERIVATION (DER)**

Rappel : Cette rubrique code, quand ils existent, les adjectifs verbaux et les dérivés nominaux de l'entrée verbale concernée : dérivés en *-able*, *-é*, *-ant*, *-age*, *-ment*, *-ion*, *-eur*, *oir(e)*, *-ure*. Les codes sont précisés pages 18-22 du livre *LVF* (version PDF).

Les correctifs apportés dans cette rubrique sont ceux indiqués dans le document daté du 11 janvier 2011 (voir plus haut la note n°3 en bas de page).

Attention : Dans le livre *LVF* (version PDF), il est écrit page 14 :

le 1er groupe indique **les adjectifs verbaux** :
en *-ant* (caractère 1)
en *-é* [*-i*, *-u*, *-t*, *-s*] (caractère 2)
en *-able* (caractère 3).

Il faut lire en fait :

le 1er groupe indique **les adjectifs verbaux** :
en *-able* (caractère 1)
en *-é* [*-i*, *-u*, *-t*, *-s*] (caractère 2)
en *-ant* (caractère 3).

Le caractère 1 code les adjectifs verbaux en *-able* (et non en *-ant*) ; le caractère 3 code les adjectifs verbaux en *-ant* (et non en *-able*).

Pour une meilleure lisibilité de ces différents types de dérivés, nous avons adjoint à la rubrique Dérivation, une colonne pour le codage de chacun de ces neuf types de dérivés : pour les adjectifs (A) verbaux : A_able, A_é, A_ant ; pour les nominaux (N) : N_age, N_ment, N_ion, N_eur, N_oir(e), N_ure.

- **Rubrique NOM (N)**

Rappel : Cette rubrique code quand il existe le mot de base, nom ou adjectif (en relation avec la rubrique Dérivation), sur lequel est formé le verbe ou code les dérivés "zéro". Les codes sont précisés page 22 du livre *LVF* (version PDF).

Aucun correctif n'a été apporté pour l'instant dans cette rubrique.

- **Rubrique LEXIQUE (L)**

Rappel : Cette rubrique code le niveau de répertoire lexicographique de l'entrée verbale concernée. Les codes sont précisés page 15 du livre *LVF* (version PDF).

Les correctifs apportés dans cette rubrique sont ceux indiqués dans le document daté du 11 janvier 2011 (voir plus haut la note n°3 en bas de page).

Pour conclure

- Le risque majeur lié à la production d'une nouvelle version est la possible introduction d'erreurs qui dégraderaient la version originale. Merci de me faire part de telles erreurs.
- Il n'est pas impossible que la réécriture en clair de certains mots abrégés m'ait échappé. Merci de me faire part de tels oublis.

Je tiendrai compte de vos remarques lors de la mise au point d'une éventuelle prochaine version.