
Michel MATHIEU-COLAS, *Les mots français à trait d'union. Problèmes de lexicographie informatique*. Thèse soutenue à l'Université Paris-Nord, le 9 mars 1993. (Le jury était composé des Professeurs Gaston Gross, Directeur, Jacques Chaurand, Président, Daniel Bresson, Maurice Gross et Robert Martin.)

Cette thèse comporte deux aspects complémentaires : d'une part la réalisation d'un dictionnaire électronique des mots à trait d'union (plus de 17000 unités lexicales sont recensées et décrites) ; d'autre part une réflexion théorique et méthodologique sur la constitution de ce dictionnaire.

Cette réflexion met elle-même en jeu deux niveaux d'analyse :

1. Il faut tenir compte en premier lieu de la **spécificité du trait d'union**. L'ambivalence formelle du signe (tout à la fois marque d'union et de séparation) reflète explicitement, sur le plan de l'écriture, l'ambiguïté constitutive de la composition. Le recensement systématique que nous avons effectué permet une analyse rigoureuse de la **morphologie** et de l'**orthographe** des unités :

typologie des formes de composition (plus de 300 types répertoriés : v. *Annexe 1*) ;
description du système des flexions (flexions terminales et flexions internes) ;
examen des variantes graphiques (soudure, disjonction, marques de nombre, etc.).

La description est complétée par une étude des locutions impliquant un trait d'union (à la *va-vite*, *boire du petit-lait*, *centre hospitalo-universitaire*) et par une présentation des principaux modèles productifs régissant les néologismes et les créations libres : composés sur particules (*l'après-68*, *post-olympique*), composés sur thèmes en -o (*pseudo-philosophie*, *franco-iranien*), formes Verbe+Nom (*protège-beauté*, *attrape-touristes*), Nom+Nom (*voyage-vacances*, *corrigé-type*), structures complexes (*la relation mère-enfant*, *une super-grande école*).

2. Au-delà des particularités liées au trait d'union, l'accent est mis sur la méthode utilisée. L'élaboration d'un **dictionnaire électronique**, conçu dans la perspective d'un traitement automatique de la langue, implique une formalisation, une représentation systématique et structurée des données lexicales. L'échantillon suivant illustre le format que nous avons adopté (voir *Annexe 2*) :

fait-tout /G:nm/M:v26/F:01/V:faitout/T:inc/C:réipient/D:cuis./R:/S:/X:A

falot-tempête /G:nm/M:nn00/F:50/V:falot tempête/T:inc/C:éclair./D:/R:région.(Suisse)/S:/X:C(2)

fan-club /G:nm/M:d62/F:60/V:/T:hum+col/C:association/D:/R:/S:/X:A;PLI92

Far-West /G:nms/M:d62/F:/V:Far West/T:inc/C:nom région/D:/R:/S:/X:RM;RC

Chaque mot-vedette fait l'objet d'un article distinct (ou "enregistrement"), décomposé en une série fixe de rubriques (ou "champs"). La barre oblique sert de séparateur, chaque champ étant étiqueté par un identificateur spécifique :

RUBRIQUES	EXEMPLE : <i>fait-tout</i>
G: catégorie grammaticale	<i>G: nm</i> (=nom masculin)
M: structure morphologique du composé	<i>M: v26</i> (=verbe + <i>tout</i>)
F: code de flexion	<i>F: 01</i> (=invariable)
V: variantes graphiques	<i>V: faitout</i>
T: traits sémantiques	<i>T: inc</i> (=inanimé concret)
C: classe d'objet	<i>C: récipient</i>
D: domaine	<i>D: cuis.</i> (=équipement de cuisine)
R: registre	<i>R:</i> (non marqué)
S: spécifications	<i>S:</i>
X: sources de la forme-entrée	<i>X: A</i> (=GDEL)

On pourra comparer notre description du mot *fait-tout* avec la présentation du *GDEL*:

"FAIT-TOUT n.m.inv. ou FAITOUT n.m. (de *fait* et *tout*). Equip. ménag. Marmite..."

fait-tout/G:nm/M:v26/F:01/V:faitout/T:inc/C:récipient/D:cuis./R:/S:/X:A

Insistons sur le fait que cette présentation linéaire, réalisable en mode ASCII, est structurellement équivalente aux deux formats traditionnels des bases de données (tableaux ou fiches), la conversion d'une forme à l'autre pouvant s'effectuer automatiquement :

TABLEAU

MOT	G:	M:	F:	V:	T:	C:	D:	R:	S:	X:
fait-tout	nm	v26	01	faitout	inc	récipient	cuis.			A
falot-tempête	nm	nn00	50	falot tempête	inc	éclair.		région. (Suisse)		C(2)
fan-club	nm	d62	60		hum+col	association				A;PLI92
Far-West	nms	d62		Far West	inc	nom région				RM;RC

FICHES

MOT :	fait-tout	MOT :	falot-tempête
CAT. GRAM. :	nm	CAT. GRAM. :	nm
MORPHOLOGIE :	v26	MORPHOLOGIE :	nn00
FLEXION :	01	FLEXION :	50
VARIANTES :	faitout	VARIANTES :	falot tempête
TRAITS :	inc	TRAITS :	inc
CLASSE :	récipient	CLASSE :	éclair.
DOMAINE :	cuis.	DOMAINE :	
REGISTRE :		REGISTRE :	région. (Suisse)
SPECIFICATIONS :		SPECIFICATIONS :	
SOURCE :	A	SOURCE :	C(2)

De quelque manière qu'elle se présente, la **formalisation** du codage facilite les opérations de traitement automatique, à commencer par les possibilités d'extraction. Veut-on connaître tous les mots du type *V + Adj* ? Il suffit de sélectionner les entrées ayant le code [M:v16] :

<i>chauffe-doux</i>	<i>passe-grand</i>	<i>porte-fort</i>
<i>coupe-net</i>	<i>passe-haut</i>	<i>sent-bon</i>
<i>gagne-petit</i>	<i>pense-petit</i>	<i>songe-creux</i>
<i>hale-bas</i>	<i>pète-sec</i>	<i>tape-dur</i>
<i>hale-haut</i>	<i>pince-dur</i>	<i>trotte-menu</i>
<i>passe-bas</i>	<i>pisse-froid</i>	

D'autre part, le format que nous proposons accorde une large place aux informations **sémantiques**, distribuées sur cinq rubriques :

- traits généraux (*humain, abstrait, etc.*)
- classes d'objets (*plantes, outils, sentiments, professions, etc.*)
- domaines (*anat., astron., mus., arbor., etc.*)
- registres (marques temporelles, régionales, sociales ou stylistiques : *fam., vx, etc.*)
- spécifications (précisions complémentaires)

Un dictionnaire électronique doit, comme tout autre dictionnaire, décrire des emplois : il en va de sa cohérence et de ses possibilités d'application (documentation, traduction, etc.). Les paramètres que nous introduisons sont de ce point de vue complémentaires. Un mot comme *centre-avant* n'est clairement déterminé que si l'on sait en même temps qu'il désigne un humain, défini par sa fonction dans le domaine du sport (football), et qu'il s'agit d'un belgicisme :

centre-avant /T:hum /C:fonction /D:footb. /R:région. (Belg.)

En particulier, la notion de **classe d'objets** (élaborée conjointement avec Gaston Gross) – ensemble de noms sémantiquement et syntaxiquement apparentés – offre de nouvelles perspectives sur la corrélation entre lexique et grammaire.

<i>jean-foutre</i>	/C: qualif.
<i>jean-le-blanc</i>	/C: oiseau
<i>jéjuno-iléite</i>	/C: inflammation
<i>jéjuno-iléon</i>	/C: part. corps
<i>jéjuno-jéjunostomie</i>	/C: acte chir.
<i>jen-min-piao</i>	/C: unité monét.
<i>jet-society</i>	/C: milieu soc.
<i>jet-stream</i>	/C: vent
<i>jette-feu</i>	/C: appar.
<i>jeu-concours</i>	/C: jeu

Un tel codage permet de rendre compte, avec beaucoup de précision, du fonctionnement syntaxique des unités : on peut TRAITER quelqu'un de *jean-foutre*, SOUFFRIR d'une *jéjuno-iléite*, FAIRE PARTIE de la *jet-society*, voir S'ENVOLER un *jean-le-blanc*, etc.

En cas d'homographie ou de polysémie, nous choisissons de **dégrouper** les entrées :

porte-serviette#1 /G:nm / ... /V: /T:inc /C:étui /D:table /R:

porte-serviette#2 /G:nm / ... /V:porte-serviettes /T:inc /C:support /D:toil. /R:

quatre-temps#1 /G:nmp / ... /V:Quatre-Temps /T:ina /C:période /D:liturg.cath. /R:

quatre-temps#2 /G:nm / ... /V:quatre temps /T:inc /C:moteur /D:moteurs /R:fam.

quatre-temps#3 /G:nm / ... /V: /T:inc /C:arbre /D:bot. /R:région.(Can.)

quatre-temps#4 /G:nm / ... /V: /T:inc /C:fruit /D:bot. /R:région.(Can.)

Chaque emploi gagne à être décrit comme une unité lexicale autonome, tant pour la clarté de la présentation que pour sa souplesse d'utilisation, dans la perspective d'un traitement automatique. Une telle décomposition, dont l'intérêt est évident du point de vue sémantique, donne en même temps plus de rigueur à la description morphologique : les exemples précédents montrent la corrélation susceptible d'apparaître entre les propriétés formelles (par ex. les variantes) et l'emploi des unités.

ANNEXE 1

EXTRAIT DE LA TYPOLOGIE : LES COMPOSÉS SUR VERBES

v00	Vb/N	<i>casse-cou</i> (N, Adj)
v01	Vb/Npropre	<i>fesse-mathieu</i> (N)
v02	Vb/Ncomp	<i>porte-cure-dent</i> (N)
v03	Dérivés de [Vb/N]	<i>tire-bouchonner</i> (V), <i>tire-bouchonné</i> (Adj), <i>tire-bouchonnement</i> (N)
v10	Vb/Déf N	<i>crève-la-faim</i> (N)
v15	Vb/Adv	<i>passe-partout</i> (N, Adj)
v16	Vb/Adj(=Adv)	<i>pète-sec</i> (N, Adj)
v20	Vb/Prép/N	<i>boute-en-train</i> (N, Adj)
v21	Vb/Prép/Déf N	<i>tape-à-l'oeil</i> (N, Adj)
v22	Vb/Prép/Adj(=N)	<i>porte-à-faux</i> (N), <i>saute-en-bas</i> (N)
v23	Vb/Prép/Divers	<i>pince-sans-rire</i> (N, Adj), <i>touche-à-tout</i> (N, Adj)
v25	Vb/Card	<i>passe-dix</i> (N)
v26	Vb/ <i>tout</i>	<i>fourre-tout</i> (N, Adj)
v27	Vb/ <i>tout</i> /Divers	<i>passe-tout-grain</i> (N), <i>laisse-tout-faire</i> (N)
v30	Vb/Vb	<i>pousse-pousse</i> (N), <i>entre-sort</i> (N)
v31	Vb/Vb/Divers	<i>cache-cache-mitoulas</i> (N)
v32	Vb/et/Vb	<i>va-et-vient</i> (N)
v33	Vb/Pprés	<i>passe-volant</i> (N)
v34	Vb/Ppass	<i>hache-bâché</i> (Adj)
v35	Vb/Vinf	<i>peut-être</i> (Adv)
v36	Vb/Divers	<i>va-nu-pieds</i> (N), <i>écoute-s'il-pleut</i> (N)
v40	Vinf/N	<i>faire-part</i> (N)
v41	Vinf/Vinf	<i>laisser-aller</i> (N)
v42	Vinf/Divers	<i>laisser-sur-place</i> (N), <i>être-au-monde</i> (N)
v50	Vimp[2sg]/Poss/N	<i>roule-ta-bosse</i> (N)
v51	Vimp[2sg]/Divers	<i>va-te-laver</i> (N)
v52	Vimp[2pl]/Pron	<i>rendez-vous</i> (N)
v53	Vimp[2pl]/Déf N	<i>cessez-le-feu</i> (N)
v54	Vimp[2pl]/Vinf	<i>laissez-passer</i> (N)
v55	Vimp[2pl]/Card	<i>chassez-huit</i> (N)
v56	Vimp[2pl]/Vimp	<i>tirez-lâchez</i> (N)
v57	Vimp[2pl]/Divers	<i>suivez-moi-jeune-homme</i> (N)
v58	Vimp[1pl]/Divers	<i>revenons-y</i> (N)
v60	Vsubj/Divers	<i>sauve-qui-peut</i> (N), <i>soit-communicé</i> (N)
v70	Pprés/N	<i>craignant-Dieu</i> (N)
v71	Pprés/Déf N	<i>battant-l'oeil</i> (N)
v72	Pprés/Divers	<i>revenant-bon</i> (N)
v80	Ppass/N	<i>cousu-main</i> (N), <i>passé-maitre</i> (N)
v81	Ppass/Prép/N	<i>laissé-pour-compte</i> (N, Adj)
v82	Ppass/Vinf	<i>oui-dire</i> (N)
v83	Ppass/Ppass	<i>coupé-cousu</i> (N, Adj)
v84	Ppass/Divers	<i>mort-flat</i> (N, Adj)
v90	V/V	<i>saisir-arrêter</i> (V)

ANNEXE 2

EXTRAIT DU DICTIONNAIRE : DEBUT DE LA LETTRE A

à-bas /G:nm/M:p57/F:01/V:/T:inc/C:lieu trav./D:ardois./R:/S:chantiers d'une ardoisière/X:A
à-côté#1 /G:nm/M:p30/F:10/V:/T:ina/C:/D:/R:/S:élément accessoire/X:A;PLI92
à-côté#2 /G:nm/M:p30/F:10/V:/T:ina/C:argent/D:/R:/S:gain ou dépense/X:A
à-coup#1 /G:nm/M:p30/F:10/V:/T:ina/C:mouvement/D:/R:/S:saccade/X:A;C;PLI92
à-coup#2 /G:nm/M:p30/F:10/V:/T:ina/C:/D:moteurs/R:/S:raté/X:A;RC;LEX
à-coup#3 /G:nm/M:p30/F:10/V:/T:ina/C:/D:/R:/S:incident/X:A;C
à-Dieu-va /G:int/M:d42/F:/V:à-Dieu-vat;à Dieu va;à Dieu vat;adieu va;.../T:/C:/D:/R:/S:/X:D
à-Dieu-vat /G:int/M:d42/F:/V:à-Dieu-va;à Dieu va;à Dieu vat;adieu va;.../T:/C:/D:/R:/S:/X:C
à-fonds /G:nmp/M:p30/F:/V:/T:ina/C:/D:/R:région.(Suisse)/S:nettoyages de printemps/X:PLI92
A-module /G:nm/M:d20/F:10/V:/T:ina/C:structure/D:algèbr./R:/S:/X:A
à-peu-près /G:nm/M:d43/F:01/V:à peu près/T:ina/C:/D:/R:/S:/X:A;PLI92
à-pic /G:nm/M:p30/F:01,10/V:/T:inc/C:form. natur./D:/R:/S:/X:A;PLI92
à-plat#1 /G:nm/M:p57/F:10/V:aplat/T:in/C:surface/D:grav.,peint.art.,impr./R:/S:teinte unie/X:A;D;PLI92
à-plat#2 /G:nm/M:p57/F:10/V:aplat/T:in/C:qualité/D:papet.,impr./R:/S:uniformité d'un papier/X:A;PLI92;GIR
à-plat#3 /G:nm/M:p57/F:10/V:/T:ina/C:mouvement/D:/R:/S:chute/X:GIR
à-propos#1 /G:nm/M:p30/F:01/V:/T:ina/C:qualité/D:/R:/S:/X:A;PLI92
à-propos#2 /G:nm/M:p30/F:01/V:/T:ina/C:oeuvre littér./D:littér./R:vx/S:/X:PLI92
à-valoir /G:nm/M:p46/F:01/V:/T:ina/C:argent/D:/R:/S:/X:A;PLI92
à-venir /G:nm/M:p46/F:01/V:avenir/T:ina/C:acte jur./D:procéd./R:/S:/X:B
abaisse-langue /G:nm/M:v00/F:01,10/V:/T:inc/C:instr./D:méd./R:/S:/X:A;PLI92
abaisse-paupière /G:nm/M:v00/F:01,10/V:/T:inc/C:instr./D:ophtalm./R:/S:/X:A
abat-carrage /G:nm/M:v03/F:10/V:/T:ina/C:opération techn./D:maroq./R:/S:/X:A
abat-carre /G:nm/M:v00/F:10/V:/T:inc/C:outil/D:maroq./R:/S:/X:A
abat-carrer /G:vt/M:v03/F:V03/V:/T:hum/C:/D:maroq./R:/S:/X:A
abat-faim /G:nm/M:v00/F:01,10/V:/T:inc/C:plat/D:alim./R:vx,fam./S:/X:C
abat-feuille /G:nm/M:v00/F:10/V:abat-feuilles/T:inc/C:dispos./D:impr./R:/S:/X:A
abat-feuilles /G:nm/M:v00/F:01/V:abat-feuille/T:inc/C:dispos./D:impr./R:/S:/X:A
abat-foin /G:nm/M:v00/F:01,10/V:/T:inc/C:ouverture,moyen d'accès/D:agr./R:/S:/X:A
abat-jour#1 /G:nm/M:v00/F:01,10/V:/T:inc/C:dispos./D:éclair./R:/S:pr lampe/X:A;PLI92
abat-jour#2 /G:nm/M:v00/F:01,10/V:/T:inc/C:ouverture/D:arch./R:/S:/X:A;D;LEX;PR
abat-jour#3 /G:nm/M:v00/F:01,10/V:/T:inc/C:dispos./D:bât./R:/S:volet incliné/X:A
abat-jour#4 /G:nm/M:v00/F:01,10/V:/T:inc/C:dispos./D:pris./R:/S:/X:A;LEX
abat-jour#5 /G:nm/M:v00/F:01,10/V:/T:inc/C:él-coiffe/D:habill./R:fam./S:visière/X:D;LEX
abat-l'oiseau /G:nm/M:v10/F:01/V:/T:ina/C:cérémonie/D:sp.(arc),coutum./R:/S:/X:B
abat-son /G:nm/M:v00/F:01,10/V:abat-sons/T:inc/C:dispos./D:arch.rel./R:/S:ds un clocher/X:A;PLI92
abat-sons /G:nm/M:v00/F:01/V:abat-son/T:inc/C:dispos./D:arch.rel./R:/S:ds un clocher/X:A;PLI92
abat-vent#1 /G:nm/M:v00/F:01,10/V:/T:inc/C:él-constr./D:bât./R:/S:auvent/X:A
abat-vent#2 /G:nm/M:v00/F:01,10/V:/T:inc/C:appar.,él-cheminée/D:chemin./R:/S:/X:A;PLI92
abat-vent#3 /G:nm/M:v00/F:01,10/V:/T:inc/C:écran,protection/D:hort./R:/S:/X:A
abat-voix /G:nm/M:v00/F:01/V:/T:inc/C:/D:arch.rel./R:/S:dessus d'une chaire/X:A;PLI92
abdomino-costal /G:adj/M:t05/F:46/V:/T:in/C:/D:anat./R:/S:/X:D(1)
abdomino-diaphragmatique /G:adj/M:t05/F:30/V:/T:in/C:/D:anat./R:/S:/X:D(1)