

# Unsupervised POS-Tagging Employing Efficient Graph Clustering

Chris Biemann, University of Leipzig, Coling/ACL SRW 2006

biem@informatik.uni-leipzig.de

## Tagset for FRENCH

This tagset was automatically induced by merely presenting large amounts of raw text to the algorithm. Numbers following hashmarks # correspond to tag labels, numbers in brackets give the number of words per tag in the lexicon. Maximally, four randomly chosen entries per tag are given. Special characters are transformed for technical reasons, e.g. `_ESENTS_` is the full stop, `_CLOSEBRAC_` is a closing bracket etc.

#1(4007):signataire, ensoleillement, sodium, archéologue #2(5927):semestriel, intimiste, exhaustif, débridé #3(3861):déviant, dilué, grand, hué #4(3986):orientait, travaillions, arrête, regardez #5(2349):réquisitionnés, endurées, ficelées, séduisants #6(4988):sacs, rhinocéros, saltimbanques, mortiers #7(3945):chorégraphie, bouderie, scénographie, sollicitation #8(2367):diverger, saisir, signifier, dissimuler #9(2164):charmants, tactiles, foraines, inquiétants, primordiaux #10(1269):FCBP, Rochebloine, Phaeton, CPLD #11(3013):sacs, syllabes, fédérations, teneurs #12(2175):titillent, demanderont, signalent, franchiront #13(2136):Teresa, Sainte-Croix, Chaussin, Présailles #14(561):programmeur, économiste, bien-aimé, officier #15(656):Khaled, Jean-Jacques, Marie-Hélène, André #16(3667):Walchhofer, Touron, Bîmes, Dahlane #17(54):Ensuite, Trop, Globalement, Physiquement #18(891):Nike, Brandt, IFPI, Science #19(4):sinon, hormis, selon, sauf #20(9):sam, Lundi, sam-dim, Mardi #21(3):saint, Saint, St #22(14):Vingt, Six, Deux, Toutes #23(46):VII, CH, N, W #24(178):niortais, minier, alsacien, stadiste #25(9):Montrer, Devant, Suivant, Faire #26(5):ceux-ci, ceux-là, celles-ci, celles-là #27(22):souhaitions, partagions, mesurons, luttons #28(8):étais, avais, aimerais, as #29(95):soixante-cinq, trois-quatre, quatre-vingt-douze, trente-et-un #30(35):seront-elles, était-ce, avait-on, avaient-ils #31(7):Toujours, Contrairement, Face, Grâce #32(5):Ma, Mon, Notre, Quelle #33(5):Troisième, Dernier, Être, Seul #34(2):ci-contre, ci-dessous #35(2):quels, quelles #36(3):peut, pouvait, pourra #37(3):sein, profit, détriment #38(2):savais, pensais #39(2):petit, nouvel #40(2):chargé, chargée #41(2):Quelles, Quels #42(2):Internet, internet #43(3):autre, immense, ultime #44(3):Je, Vous, Tu #45(3):Hauts-de-Seine, Yvelines, Bouches-du-Rhône #46(2):Si, Lorsque #47(2):tient, tenait #48(2):cependant, toutefois #49(3):En, Fin, Début #50(2):mon, ma #51(2):celui-ci, celle-ci #52(2):seront, furent #53(2):olympiques, Olympiques #54(2):sait, savait #55(7):Beaucoup, susceptibles, Afin, susceptible #56(171):politiques, américains, publics, sociaux #57(2):hésite, hésitent #58(3):Saddam, Vivendi, sifflet #59(4):mais, voire, sinon, tantôt #60(2):nouvelle, petite #61(2):favorable, hostile #62(2):techniques, primaires #63(2):espérons, pensons #64(2):dont, puisque #65(3):Van, Maria, Di #66(3):pas, guère, nullement #67(3):grands, grandes, futurs #68(165):Lyon, Paris, Saint-Étienne, Bagdad #69(3):souvent, félicite, réjouit #70(2):très, relativement #71(2):vainqueur, fondateur #72(2):longue, courte #73(2):abord, emblée #74(2):Vendredi, Jeudi #75(3):choisi, refusé, accepté #76(2):Ses, Leurs #77(2):Outre, Concernant #78(2):féminine, masculine #79(9):septembre, juin, janvier, avril #80(7):explique-t-il, ajoute-t-il, poursuit-il, affirme-t-il #81(2):nuit, continuité #82(3):négociations, discussions, pourparlers #83(2):Canal+, Microsoft #84(2):forcément, nécessairement #85(10):10e, 13e, 11e, 18e #86(2):âgées, handicapées #87(2):grâce, appartenant #88(2):bonne, mauvaise #89(2):faisons, ferons #90(2):appartements, exploitations #91(8):son, sa, ses, leur #92(2):Ça, Ca #93(2):Ils, Elles #94(2):expliquer, illustrer #95(2):républicain, libéral #96(3):York, Angeles, Aires #97(2):tard, tôt #98(2):contrôle, commandement #99(2):petits, petites #100(2):maire, député-maire #101(2):nucléaire, baroque #102(2):importants, importantes #103(2):premiers, premières #104(2):semble, semblaient #105(2):murs, ruines #106(2):relations, liens #107(2):êtres, ordures #108(2):rails, épaules #109(2):installé, installée #110(2):soir, matin #111(2):metteur, rédacteur #112(2):présidente, directrice #113(2):défendre, protéger #114(4):vingt, trente, cinquante, quarante #115(2):eux-mêmes, elles-mêmes #116(3):capables, obligés, incapables #117(2):matches, matchs #118(3):peux, vois, pouvais #119(2):tous, toutes #120(5):oeuvre, examen, \_#156\_uvre, otage #121(2):Télécom, Télévisions #122(7):%, km, kg, milles #123(2):inscrit, inscrivent #124(2):Comité, Fonds #125(2):engagements, obligations #126(2):Seuls, Seules #127(2):réalisé, effectué #128(2):Bellecour, Carnot #129(3):mars, mai, février #130(2):ils, elles #131(2):gauche, droite #132(2):pratiques, traitements #133(2):repères, dettes #134(2):trop, assez #135(4):permis, décidé, tenté, écopé #136(2):contrairement, conformément #137(2):prêt, prête #138(2):attendons, demandons #139(3):responsables, représentants, défenseurs #140(2):Tous, Toutes #141(32):voitures, soins, violences, médicaments #142(2):prend, prenait #143(4):dans, sur, sous, devant #144(2):lycée, collège #145(4):autour, Lors, au-dessus, Autour #146(2):parquet, rectorat #147(3):telle, certaine, brève #148(3):saoudite, latine, potable #149(2):issus, issues #150(2):parle, parlait #151(2):prévue, programmée #152(2):travailler, rouler #153(2):déclenchement, rétablissement #154(3):notamment, essentiellement, principalement

*see next page*

#155(2):tels, telles #156(6):vigueur, lice, vain, amont, péril, suspens #157(2):risque, taux #158(2):heures, degrés #159(14):socialistes, industriels, parlementaires, communistes  
 #160(2):totalement, complètement #161(2):certain, joli #162(2):su, semblé #163(2):salon, rallye #164(4):Par, Très, Côté, DU #165(4):doit, devrait, pourrait, devait #166(2):Laporte,  
 Choulet #167(2):économiste, animatrice #168(2):veux, voulais #169(2):il, elle #170(2):également, néanmoins #171(2):spécialisés, spécialisées #172(2):agents, officiers  
 #173(2):Gironde, NBA #174(5):vont, doivent, devraient, devront, devaient #175(2):gain, tremblement #176(2):né, née #177(3):reçu, obtenu, subi #178(2):veut, voulait  
 #179(2):propose, proposait #180(5):qu, S, lorsqu, puisqu, Lorsqu #181(2):chemin, contenu #182(2):doute, relâche #183(2):responsable, spécialiste #184(3):étrangères, unies,  
 Unies #185(2):certains, certaines #186(2):Celui-ci, Celle-ci #187(22):déclaré, annoncé, indiqué, affirmé #188(2):centaines, dizaines #189(4):New, San, Los, Buenos  
 #190(13):arrêter, occuper, appliquer, attaquer #191(2):marqué, dominé #192(2):devenu, devenue #193(2):changement, parfum #194(3):président, procureur, préfet  
 #195(3):peuvent, pourront, pouvaient #196(2):commencé, continué #197(2):rival, coéquipier #198(3):âgé, âgés, âgée #199(2):italien, espagnol #200(3):trouve, retrouve, retrouvait  
 #1(4):plupart, totalité, multiplication, quasi-totalité #202(2):solitaire, solo #203(4):faut, faudra, fallait, faudrait #204(3):a, avait, aurait #205(2):excellence, intérim #206(2):Figaro,  
 Parisien #207(2):Près, Agé #208(2):directeur, coordinateur #209(2):rappelle, rappelait #210(2):étrangers, intermédiaires #211(3):tourisme, pétrole, tramway #212(2):secteur,  
 paysage #213(2):Tête, Légion #214(2):visiteurs, participants #215(2):adaptés, adaptées #216(2):américain, brésilien #217(2):prévisions, préparatifs #218(2):Parlement, parlement  
 #219(2):Ceux-ci, Celles-ci #220(2):jusqu, Jusqu #221(5):millions, milliards, million, milliard, centimes #222(2):besoins, exigences #223(2):peuple, capitalisme #224(2):largement,  
 fortement #225(2):celui, celle #226(2):départements, disciplines #227(2):aucun, aucune #228(2):Pen, Guen #229(2):Sri, Moyen #230(2):différents, différentes #231(2):est, était  
 #232(2):Quelque, Environ #233(2):voit, voyait #234(2):FC, Secours #235(2):sais, comprends #236(2):Juninho, TPS #237(2):sud, nord #238(5):Talaudière, Rochelle, Haye,  
 Fouillouse #239(2):A, À #240(5):permet, permettra, permettent, permettrait #241(2):Au, Aux #242(2):utilisant, accueillant #243(2):situé, située #244(2):font, faisaient  
 #245(2):restent, restaient #246(2):aider, inciter #247(2):nom, siège #248(2):veulent, voulaient #249(2):revient, revenait #250(2):prêts, prêtes #251(2):nombreux, nombreuses  
 #252(2):bande, baie #253(2):député, sénateur #254(2):perdu, gagné #255(2):le, les #256(3):Dominique, Gilles, Geneviève #257(3):possibilité, nécessité, notion #258(2):vit, vivait  
 #259(2):pense, crois #260(3):Son, Sa, Leur #261(2):Me, Mgr #262(2):concernés, concernées #263(2):minutes, secondes #264(2):pris, repris #265(4):national, régional, municipal,  
 central #266(2):divers, diverses #267(4):E, m2, hectares, E\_ #268(2):rencontre, confrontation #269(3):ont, avaient, auraient #270(2):Tandis, Rappelons #271(2):auprès, vis-à-vis  
 #272(2):hip-hop, reggae #273(3):parents, créations, fournisseurs #274(2):vote, dépistage #275(2):proposer, fournir #276(2):appelle, appelait #277(23):obtenir, ouvrir, apporter,  
 organiser #278(7):samedi, dimanche, vendredi, mardi #279(2):Mais, Car #280(2):Non, Oui #281(4):suis, vais, dois, serai #282(3):coeur, c\_&156\_ur, cur #283(3):victoire, défaite,  
 condamnation #284(2):seuls, seules #285(2):comité, groupement #286(2):Créée, Créé #287(2):professionnel, naturel #288(2):camp, coffre #289(2):document, phénomène  
 #290(2):cette, cet #291(2):et, ou #292(2):concert, stage #293(2):Champions, sceaux #294(2):semble, semblait #295(3):étions, aurons, aurions #296(2):aujourd, Aujourd  
 #297(2):un, une #298(3):parcours, modèle, socle #299(2):retraite, vitesse #300(2):prévoit, prévoyait #301(1):ne #302(1):cas #303(1):beaucoup #304(1):De #305(1):Les  
 #306(1):y #307(1):\_OPENBRAC\_ #308(1):ans #309(1):faire #310(1):après #311(1):mois #312(1):politique #313(1):lors #314(1):qui #315(1):étaient #316(1):Et #317(1):Cette  
 #318(1):première #319(1):contre #320(1):\_ESENTE\_ #321(1):va #322(1):\_NUMBER\_ #323(1):il #324(1):- #325(1):la #326(1):ministre #327(1):C #328(1):années #329(1):pays  
 #330(1):pour #331(1):temps #332(1):bien #333(1):encore #334(1):se #335(1):Pour #336(1):où #337(1):Une #338(1):aux #339(1):être #340(1):général #341(1):fin #342(1):avoir  
 #343(1):si #344(1):ces #345(1):depuis #346(1):Dans #347(1):c #348(1):hier #349(1):que #350(1):dit #351(1):des #352(1):par #353(1):soit #354(1):même #355(1):jours  
 #356(1):aussi #357(1):au #358(1):public #359(1):nouveau #360(1):quelques #361(1):Nous #362(1):personnes #363(1):peu #364(1):nous #365(1):plusieurs #366(1):du  
 #367(1):On #368(1):Des #369(1):on #370(1):Un #371(1):été #372(1):hui #373(1):moins #374(1):année #375(1):\_ESENTS\_ #376(1):\_KOM\_ #377(1):\_CLOSEBRAC\_  
 #378(1):ainsi #379(1):premier #380(1):avant #381(1):\_APOS\_ #382(1):sont #383(1):alors #384(1):donc #385(1):eu #386(1):tout #387(1):cours #388(1):déjà #389(1):Le  
 #390(1):deux #391(1):\_ #392(1):entre #393(1):plus #394(1):jour #395(1):toute #396(1):lui #397(1):non #398(1):toujours #399(1):\_ESENTQ\_ #400(1):sera #401(1):je #402(1):j  
 #403(1):comme #404(1):grand #405(1):Elle #406(1):euros #407(1):en #408(1):de #409(1):fois #410(1):dernier #411(1):place #412(1):ce #413(1):\_QUOT\_ #414(1):partie  
 #415(1):La #416(1):\_COL\_ #417(1):sans #418(1):Ce #419(1):fait #420(1):à #421(1):avec