

Unsupervised POS-Tagging Employing Efficient Graph Clustering

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

biem@informatik.uni-leipzig.de

Tagset for ITALIAN

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(3949):cumulo, Pubblico, nervosismo, distaccamento #2(5165):osserviamo, includa, perseguitano, studiavo #3(3157):insufficienza, schiarita, ringhiera, pendenza #4(4207):polverizzato, risolutamente, sorprendentemente, resasi #5(4366):silenzi, veltroniani, lanci, sciocchi #6(1620):necropoli, salumeria, Frattese, Montefibre #7(3591):scontrare, intrattenere, digitare, svilupparne #8(2780):caramelle, bande, sopraffazioni, femmine #9(8013):Rivelli, Buozzi, Zaccaria, Puccini #10(4244):insoluti, indotte, risucchiati, uguali #11(2616):Kandahar, Oriente, Podgorica, Monteverde #12(1208):architettonico, distributivo, statistico, misto #13(1190):recante, unitaria, giallazzurra, infantile #14(283):Paura, Inoltre, Meno, Vanno #15(2985):emblematica, superficiale, celebre, berlusconiana #16(811):smantellando, spulciando, analizzando, consumando #17(653):ergastolano, geometra, musicologo, ventitreenne #18(567):londata, interpartitico, lalba, linconveniente #19(162):vogliano, saprete, devo, preferiremmo #20(563):nazionaliste, chirurgiche, giornaliere, aeroportuali #21(308):settecentomila, Quindici, Torchiagina, 3mila #22(261):contributivi, automatici, epici, pedagogici #23(58):venisse, avessi, abbiamo, fui, #24(620):Teresa, Amalia, Martina, Edmondo #25(121):papà, Padre, vicesindaco, vice-sindaco #26(52):Dobbiamo, Mi, Posso, Bisogna #27(48):canonica, settima, Terza, tredicesima #28(488):napoletani, nolani, soccorsi, supporters #29(76):abbastanza, straordinariamente, altrettanto, fondamentalmente #30(34):dellassise, dellamministrazione, dellassociazione, sulleconomia #31(15):sebbene, lAst, essa, linnovazione, esso #32(8):Martedì, Giovedì, Lunedì, Leuro #33(52):farla, potervi, poterla, volerli #34(2):bin, Bin #35(2):dellufficio, dellUfficio #36(6):inutilmente, dallAscoli, stancamente, invano #37(236):Teresa, Guido, Max, Angela #38(8):lUnione, lassociazione, lambasciata, lAssociazione #39(31):occidentale, avellinese, milanese, ovest #40(38):Altre, Diciotto, Sei, Sette #41(6):signor, mio, suo, nostro #42(8):Gasperi, Mita, Crescenzo, Canio #43(2):affinchè, affinché #44(5):dellUniversità, delluniversità, alluniversità, allUniversità #45(8):NOTE, Arbitro, Ammoniti, Arbitri #46(12):manciata, cinquantina, bottiglia, quarantina #47(54):cosidetto, ventunesimo, terzultimo, 14esimo #48(10):avermi, averla, averlo, essermi #49(2):statale, Statale #50(16):contestualmente, stamane, nellanticipo, allantistadio #51(45):lesponente, lospite, loncologo, leurodeputato #52(5):moltissimo, pochissimo, sodo, parecchio #53(26):unintesa, lindagine, loperazione, unispezione #54(6):dellattuale, dellantico, nellex, dellallora #55(6):-aggiunge, -spiega, -afferma, -è #56(3):destra, sud-occidentale, sinistro #57(24):gospel, country, funk, reggae #58(2):cè, cera #59(15):avere, ottenere, aprire, effettuare #60(2):Democratici, tassi #61(2):giornata, seduta #62(2):curva, pianta #63(12):presentati, registrati, svolte, ritrovati #64(3):L, All, Nell #65(4):città, stagione, regione, cittadina #66(2):musicale, teatrale #67(2):siamo, saremmo #68(2):ricerca, corte #69(3):nemmeno, neanche, neppure #70(2):donna, ragazza #71(3):economica, fisica, estera #72(4):tratta, trattava, tratterà, tratterebbe #73(2):partecipato, aderito #74(2):ASCOLI, MACERATA #75(12):settembre, gennaio, giugno, luglio #76(3):rappresentanti, responsabili, promotori #77(2):finora, sinora #78(4):hanno, avevano, avrebbero, abbiano #79(2):movimento, debito #80(3):squadre, formazioni, compagini #81(2):al, nel #82(6):possibilità, presenza, richiesta, necessità #83(2):Perché, Perché #84(2):riguarda, concerne #85(2):Castel, Casal #86(13):annunciato, confermato, ricordato, sostenuto #87(2):bloccato, sorpreso #88(4):Queste, Tutte, Molte, Tante #89(2):dà, dava #90(2):Grottaminarda, Lioni #91(7):nuova, piccola, vecchia, splendida #92(2):accesso, iscrizione #93(2):spettacolo, schieramento #94(2):annuncia, rivela #95(2):tiene, tengono #96(2):davanti, dinanzi #97(2):Reti, RETI #98(2):nulla, niente #99(2):Né, Nè #100(2):Liniziativa, Loperazione #101(2):organizzato, promosso #102(13):il, dell, all, dall, nell, D, sull, Sant, quell, anch, Dell, piu #103(2):dichiarazioni, affermazioni #104(2):mai, affatto #105(2):Nessun, Nessuna #106(2):paesi, Paesi #107(2):Internet, internet #108(1):volley #109(5):il, la, i, le, gli #110(2):amministrazione, Amministrazione #111(2):Ha, Hanno #112(2):sa, capisce #113(3):Sì, No, Beh #114(2):definitivamente, ulteriormente #115(2):Unaltra, Altra #116(2):dellInterno, dellAmbiente #117(5):Alle, under, Ore, Under #118(8):sua, nostra, stessa, mia #119(3):direttore, segretario, coordinatore #120(2):comincia, comincerà #121(3):bisogna, dobbiamo, dovremo #122(3):Sembra, Pare, Possibile #123(2):quest, Quest #124(2):perché, perchè #125(2):sinistra, destra #126(2):Serve, Esiste #127(2):polemica, trattativa #128(2):Calabria, Sabina #129(2):Corte, Conferenza #130(17):realizzazione, raccolta, costruzione, firma #131(2):Alema, Ambrosio #132(2):morte, produzione #133(2):Con, Attraverso #134(12):napoletano, tedesco, russo, reatino #135(2):tra, fra #136(6):dare, trovare, creare, diventare #137(2):intenzione, ondata #138(2):Lui, Questultimo #139(3):Ora, Adesso, Stavolta #140(2):confronto, dialogo #141(2):dellintera, dellintero #142(2):ormai, oramai #143(2):campi, tornei #144(2):vigili, Vigili #145(2):scelto, accettato #146(2):ovunque, dappertutto #147(2):arrivano, arriveranno #148(3):crisi, crescita, guardia #149(2):bella, brutta #150(2):dallamministrazione, allamministrazione #151(3):ha, aveva, avrebbe #152(2):torna, tornerà #153(2):servizio, concorso #154(2):Dice, Spiega #155(2):organizzata, promossa #156(2):Domani, Stasera #157(2):commissariato, Commissariato #158(2):soltanto, solamente #159(3):vengono, verranno, venivano #160(2):Procura, procura #161(2):lui, lei #162(2):buone, ottime #163(2):alcuni, alcune #164(2):Durante, Presso #165(2):sembrano, risultano #166(1):urbana #167(2):Stessa, Stesso #168(1):aiutare #169(2):resta, rimane #170(2):sta, stanno #171(2):riusciti, riuscite #172(5):Sessanta, Settanta, Ottanta, Trenta #173(2):col, coi

see next page

#174(2):evitare, impedire #175(2):lunico, lunica #176(3):Stati, enti, Enti #177(2):lascia, lasciano #178(2):lanciato, firmato #179(2):sono, sarebbero #180(21):dice, spiega, afferma, aggiunge
 #181(2):realizzato, costruito #182(2):Tra, Fra #183(4):possono, potrà, poteva, potevano #184(2):noi, voi #185(2):va, andava #186(2):culturale, spirituale #187(2):cercando, tentando #188(2):partito,
 malato #189(2):rosse, vuote #190(3):Forza, Miss, Sviluppo #191(2):fino, sino #192(2):Il, La #193(2):Luca, Simone #194(6):Nella, Sulla, Ai, Sulle, Agli, Sui #195(2):tv, Tv #196(2):salvezza,
 retrocessione #197(2):davvero, veramente #198(2):ministero, Ministero #199(2):proprietaria, vincitrice #200(3):viene, verrà, veniva #201(2):Ponticelli, Fuorigrotta #202(2):Coppa, coppa #203(2):nord,
 sud #204(2):miele, Casagiove #205(2):aumento, incremento #206(2):conclusioni, testimonianza #207(2):nata, avvenuta #208(2):stava, stavano #209(2):alcun, alcuna #210(2):ultime, prossime
 #211(144):fare, vincere, prendere, raggiungere #212(2):minuti, secondi #213(2):Centro, Circolo #214(2):Altri, Altre #215(2):presentato, introdotto #216(2):silenzio, paradiso #217(3):professor, dottor,
 prof_ #218(2):palla, sfera #219(5):del, della, dei, delle, degli #220(2):svoltasi, svoltosi #221(2):scopre, augura #222(2):All_, Allenatore #223(3):M_, G_, F_ #224(2):voluto, saputo #225(2):informatica,
 Alitalia #226(19):nazionale, sociale, internazionale, popolare #227(2):pratica, sostanza #228(2):femminile, maschile #229(5):all'interno, nell'ambito, all'altezza, all'insegna, all'ingresso #230(2):positivo,
 negativo #231(2):mercato, federalismo #232(2):alte, basse #233(2):incontro, accordo #234(19):detto, spiegato, dichiarato, aggiunto #235(2):tutte, entrambe #236(2):illegale, abusivo #237(2):ingresso,
 esordio #238(3):Vetere, all'aperto, dell'Umbria #239(2):pubblici, archeologici #240(3):mondo, settore, comparto #241(2):presso, durante #242(2):Valentino, Sebastiano #243(6):trasferita, piena, materia,
 funzione #244(2):cercare, tentare #245(2):parla, parlava #246(2):montana, Montana #247(2):Pietro, Biagio #248(2):completamente, totalmente #249(2):sanno, sapevano #250(4):Comune, comune,
 Tribunale, tribunale #251(2):un, una #252(2):originario, originaria #253(2):coordinati, coordinata #254(8):ROMA, PESCARA, MILANO, LANCIANO #255(2):maglia, casacca #256(2):vertici, cancelli
 #257(2):pieni, piene #258(15):giocare, lavorare, vivere, crescere #259(3):municipale, Civile, Municipale #260(2):l, Le #261(2):Pubblica, Protezione #262(2):milione, miliardo #263(2):Dall, Sull
 #264(2):riforma, liberalizzazione #265(31):politici, popolari, romani, napoletani #266(2):altrettanti, altrettante #267(3):buon, grosso, brutto #268(2):impegnato, impegnata #269(2):ampi, ampie
 #270(4):vuole, vogliono, voleva, volevano #271(2):prevede, prevedeva #272(4):c, C, n, com #273(3):vecchio, doppio, povero #274(2):d'Italia, d'Europa #275(2):preda, sella #276(2):superiore, inferiore
 #277(2):avevo, avevamo #278(4):Giorgio, Lorenzo, Giacomo, Donato #279(2):avrà, ebbe #280(2):momenti, periodi #281(2):a, in #282(18):Ancona, Avellino, Ascoli, Aversa #283(2):portato, spinto
 #284(14):trasferimento, profilo, cammino, consenso #285(2):riuscito, riuscita #286(2):Qui, Spesso #287(2):nello, dallo #288(2):macchina, barca #289(2):avendo, essendo #290(3):attesa, virtù,
 prossimità #291(2):calendario, cartellone #292(2):nessuna, nessun #293(2):of, and #294(3):insieme, accanto, assieme #295(2):l'intera, l'intero #296(2):Michelangelo, Martin #297(13):alla, nella, ai,
 dalla #298(4):causa, favore, bordo, ridosso #299(2):tutt'altro, colui #300(2):poter, potersi #301(3):partire, giudicare, prescindere #302(2):far, fargli #303(14):gruppo, ruolo, colpo, clima #304(2):via,
 viale #305(5):potrebbe, deve, dovrebbe, dovrà #306(22):gestione, chiusura, presentazione, consegna #307(2):territorio, litorale #308(4):disposizione, tavolino, sedere, galla #309(2):qual, Qual
 #310(2):Pro, Polisportiva #311(2):vanno, andavano #312(2):arriva, arrivava #313(4):regionale, provinciale, palestinese, federale #314(2):allenarsi, hoc #315(5):domenica, sabato, giovedì, martedì,
 mercoledì #316(2):estate, inverno #317(2):quest'ultimo, quest'ultima #318(2):vertice, comando #319(2):restano, rimangono #320(201):cose, donne, indagini, strade #321(4):è, era, sarebbe, é
 #322(2):sicuramente, certamente #323(2):Fiamme, Nazioni #324(2):grado, possesso #325(2):fondamentale, essenziale #326(4):per, con, attraverso, tramite #327(2):confronti, pressì #328(1):ne
 #329(1):comunale #330(1):Non #331(1):come #332(1):da #333(1):Per #334(1):_OPENBRAC_ #335(1):ci #336(1):fine #337(1):senza #338(1):meno #339(1):e #340(1):secondo #341(1):parte
 #342(1):alle #343(1):saranno #344(1):In #345(1):tutto #346(1):qualche #347(1):Nel #348(1):_NUMBER_ #349(1):chi #350(1):uno #351(1):Ma #352(1):- #353(1):proprio #354(1):anno #355(1):mi
 #356(1):nuovo #357(1):fa #358(1):se #359(1):già #360(1):grande #361(1):oggi #362(1):loro #363(1):cui #364(1):primo #365(1):si #366(1):fatto #367(1):anni #368(1):anche #369(1):casa
 #370(1):sia #371(1):questi #372(1):dalle #373(1):sempre #374(1):posto #375(1):corso #376(1):giorni #377(1):o #378(1):su #379(1):quanto #380(1):contro #381(1):E #382(1):Un #383(1):lo
 #384(1):Se #385(1):essere #386(1):momento #387(1):punto #388(1):Si #389(1):_ESENTS_ #390(1):_KOM_ #391(1):Sono #392(1):dello #393(1):questa #394(1):_CLOSEBRAC_ #395(1):volta
 #396(1):poco #397(1):stesso #398(1):_APOS_ #399(1):Una #400(1):quale #401(1):di #402(1):caso #403(1):ore #404(1):soprattutto #405(1):solo #406(1):tempo #407(1):quello #408(1):_
 #409(1):non #410(1):nei #411(1):_ESENTQ_ #412(1):prima #413(1):ora #414(1):oltre #415(1):giorno #416(1):ieri #417(1):quella #418(1):ogni #419(1):ed #420(1):Gli #421(1):così #422(1):ancora
 #423(1):sarà #424(1):questo #425(1):erano #426(1):_QUOT_ #427(1):ad #428(1):più #429(1):A #430(1):che #431(1):_COL_ #432(1):Anche #433(1):dal #434(1):tutti #435(1):altri #436(1):bene