deep parsing watson m. c. mccord j. w. murdock b. k. boguraev deep parsing component english slot grammar esg parser predicate argument structure pas builder provide core linguistic analysis question text content component fundamental question analysis candidate generation analysis passage evidence watson project esg enhance performance jeopardy!i question established reference datum improve pas build esg support high level analytic paper describe component illustrate pattern base relation extraction component watson provide quantitative result evaluate component level performance esg parsing deep parsing component english slot grammar esg parser predicate argument structure pas builder provide core linguistic analysis question hypothesize answer specifically component fundamental question analysis candidate generation analysis passage evidence 1–3 esg 4–7 deep parser sense parse tree produce sentence segment phrasal category level logical analysis deep structure parse tree show surface level grammatical structure surface structure deep structure parse tree segment rank accord parse scoring system describe watson use highest rank parse parse score roughly correspond likelihood parse correct paper provide overview slot grammar sg current state discuss new feature special adaptation jeopardy question answer qa task improvement motivate jeopardy challenge applicable general english application adaptation special jeopardy question control flag setting default turn esg jeopardy task parse analysis esg follow application pas builder simplifie abstract esg parse variety way example drop term e.g. auxiliary verb rarely important task downstream component perform active passive alternation

bjohn sell fish ba fish sell john slightly different structure esg structure pas deep parsing suite watson consist esg follow pas builder deep parsing result pervasively watson qa system component stage deepqa architecture 8 question analysis question decomposition hypothesis generation hypothesis evidence scoring etc specific relation extraction 9 section relation extraction identifies semantic relationship sense section entity result deep parsing question analysis 1 use result deep parsing identify type answer question keyword search component 2

use semantic relation question identify keyword strong semantic connection question ask keyword give high weight search query passage score component 3 use result deep parsing question text passage find keyword search determine passage align question provide evidence support candidate answer royalty free permission computer base information service system permission republish portion paper obtain editor m. c. mccord et al 3 1 type coercion component 10 use pas compare type request answer type find result pas relation extraction large corpus aggregate prismatic knowledge base 11 variety search 2 answer scoring 3 10 component pervasive usage deep parsing result reflect fact component provide core natural language processing capability watson qa system paper first describe sg parsing watson discuss pas follow approach relation extraction illustrate use esg pas main section follow section use watson evaluation related work conclusion future work sg parsing system divide large language universal shell language specific grammar english german french spanish italian portuguese sg feature describe section shell specific english esg example esg discuss 1 pipeline sg parsing 2 nature sg parse 3 lexical system 4 syntactic analysis end paper describe evaluation esg performance pipeline sg parsing main step sg parsing tokenization segmentation b morpholexical analysis c syntactic analysis step self contain handle tagging system html output directly component watson unlike parser sg use speech pos tagger correspond information simply come syntactic analysis follow

speech pos tagger correspond information simply come syntactic analysis follow describe nature sg parse tree concentrate lexicon syntactic analysis nature sg analysis figure 1 show sample jeopardy question esg parse tree look example overview sg parse general sg parse tree dependency tree tree node n center headword surround left right modifier turn tree node modifier m n fill slot n. slot show grammatical role m n. example node headword bchandelier fill subj i.e. subject slot coordinated vp verb phrase node headword bbut modifier tree structure surface structure parse analysis sample parse display left surface structure tree linesveach line connect node m mother node n show slot fille m n. slot kind complement slot

adjunct slot complement slot subj obj i.e. direct object verb idiosyncratic sense headword associate sense lexical entry adjunct slot vadv verb modify adverbial associate pos headword sense sg syntax module adjunct slot-filler esg parse jeopardy question bchandelier look great nowadays usually use item derive

m. c. mccord et al typically modify node category pos associate complement slot play dual role sg grammatical role mention addition logical argument word sense describe later section sort parse display give figure 1 line row correspond 1 to-1 tree node describe five main ingredient associate parse node state part constitute deep structure constitute surface structure 1 headword nodevthe internal parse datum structure store version headword include form occur text inflected mixed case etc b lemma citation form c sg word sense node explain bsg lexicon subsection typically headword come single word token multiword punctuation symbol act coordinator special symbol special kind node quote node describe form parse display headword show lemma form display option form headword see middle column predicate follow argument example derive(17 u 15 12 predication word sense predication node main vehicle show deep structure parse 2 id nodevthis integer case word number headword segment exception common multiword id word number head multiword parse display node id show first argument word sense predication example 17 derive(17 u 15 12 3 logical deep argument frame nodevin internal parse datum structure consist list complement slot word sense slot associate filler node nil filler bderive node example list pair subj nil obj ph15 comp ph12 ph15 phrase node id 15 span btheir ph12 phrase id 12 span bfrom subj slot overt filler note bderive give passive slot-filler constitute logical active form argument verb example ph15 logical obj bderive grammatically subj bbe speak logical deep argument frame node

verb first member argument frame logical subject word sense predication node form parse display predicate word sense optionally citation form first argument node id remain argument id filler node argument frame u bunfilled bunknown word sense predication directly translate

logical predication replace numerical argument similarly index logical variable example deriveðe17 x x15 x12þ general deriveðe x y zþ mean e event x derive y manner z. node id argument think event argument generally entity argument predication note example bchandeliers node 1 show logical subj predicate blook bdo buse surface structure role grammatical subj coordinate node 4 handle coordination sg parsing algorithm bfactor slot conjunct happen node 2 6 provide common subj fille node 1 show 1 logical subj conjunct sg parsing fill implicit argument nonfinite vp result 1 logical subj node 9 sample parse show kind implicit argument fille predication great(3 1 u bchandelier 1 fill first slot asubj adjective bgreat directly show bgreat apply bchandelier context blook b predication which(13 11 u bitem 11 fill first slot nsubj relative pronoun bwhich interpret show relative pronoun co referent bitems build logical form relative pronoun variable simply replace variable bitems 4 feature nodevin parse display node feature list right headword morphosyntactic semantic feature example feature omit brevity sake first feature list pos node feature come headword sense obtain morpholexical analysis headword add syntactic analysis node acquire modifier 5 surface modifier structure nodevin internal parse datum structure node n associate listsvfor left modifier premodifier right modifier postmodifier modifier node pair slot fill n. parse display m. c. mccord et al 3 3

clear read tree structure line dot left display picture tree diagram standard form turn slot show close headword node n slot n fill mother node complement slot s slot option s show parenthesis s. instance node 17 bderived fill slot pred(en mean node 17 fill past participial form pred predicate slotvfor bbe node 16 information slot option give subsection core sg deep structure set word sense predication describe 3 close logical predication deep structure information exist semantic feature node morphosyntactic feature tense number matter logical form core surface structure lie headword information 1 2 morphosyntactic feature surface modifier structure 5 adjunct slot appear 5 relevance deep structure e.g. determiner produce quantifier logical form subsection describe sg lexical system improvement benefit watson sg

analysis process drive lexicon particularly sg lexicon specify complement slot frame word sense main step syntactic analysis slot-filling sg lexical entry typically index citation form word multiword morpholexical analysis token efficient lookup lexicon morphological analysis inflectional derivational esg morphology currently handle 29 derivational affixe language version sg esg main lexicon call base lexicon system allow number lexicon one base lexicon typically user addendum lexicon esg base lexicon approximately 87,000 entry word form recognize derivational inflectional morphology work watson develop way augment i.e. expand improve esg base lexicon automatically source particularly princeton wordnet 12 13 process augmentation run time new version base lexicon take 5 second

standard desktop describe augmentation method following describe form sg lexical entry b improvement work form sg lexical entry follow sample entry slightly simplified esg base lexicon talk g v obj n p comp p g v obj1 comp1 p g n nsubj nobj n p ncomp p general lexical entry index word give citation lemma form single word multiwordvtalk example follow sequence sense frame wordvthree example verb frame noun frame sense frame specify following seven kind item optional first 1 pos 2 complement slot frame 3 featuresvboth semantic syntacticv(4 word sense 5 numerical score 6 subject area test 7 generalized support verb construction sample show 1 2 sense frame sense frame define sg word sense index word esg lexical word sense syntactic nature differ slot frame constrain possible semantic word sense sg framework allow finer semantic distinction word sense slot option semantic type test slot filler extent esg lexicon let look detail seven kind item sense frame speech parse data structure 15 possible part speech include noun verb adj adv qual qualifier det prep subconj subordinate conjunction conj coordinate conjunction see sample parse tree figure 1 pos list first node feature lexicon use pos name case noun verb sake brevity instance lexical pos n expand noun cn common noun parse tree v expand verb feature number tense add basis morphology complement slot frame sample entry first sense frame talk show slot obj comp slot frame initial subj slot imply verb subj slot omit long need special option obj slot show

option n p

first allow np noun phrase filler obj plus nominal gerund phrase second allow babout[-pps prepositional phrase option disjunctively view comp slot option m. c. mccord et al p allow bto[-bwith[-pps slot frame allow variant bjohn talk mathematic bill general slot associate list possible slot option symbol constraint possible filler slot apply adjunct complement slot option adjunct slot specifie slot syntax module idea complement slot slot deal argument word sense realize different way syntactically slot option specify basic syntactic category filler kind test filler semantic type requirement feature test subject area test test specific word recursively boolean combination test example show lexical entry specific option complement slot specified obj1 slot second sense frame talk default option assign system slot optional default mean require fille valid use sense frame apply slot list first sense frame example suffix 1 second sense frame indicate slot obligatoryvthat fille feature syntactic feature semantic type esg currently use approximately 160 semantic type belong type hierarchy example see sample parse figure 1 e.g. artf artifact langunit language unit type mainly high level include e.g. physical object substance abstraction property natural phenomenon event act change type time location measure collection live human nonhuman animal communication feeling profession artifact artistic composition important type effect parsing brole entity[van entity usually person view have particular role bteacher bleader bmother etc detail 7 feature appear sense frame transfer word phrase have word sense head bstarter phrase syntactic analysis esg syntax rule test omit word sense take index word entry suitable integer append current sense frame convert bstarter phrase parsing item word sense node mention precede subsection syntax rule test specific sense name word feed parse scoring system describe reward penalize use current sense subject area test subject area domain topic current text specifie determine automatically e.g. computer medicine kind test allow exclude

specifie determine automatically e.g. computer medicine kind test allow exclude current sense

numerical score generalized support verb construction bmake reference noun breference idiosyncratically use support verb bmake special keyword allow store information noun breference overload common verb bmake mechanism general handle multiword allow variation intersperse modifier extraposition noun bwh question mechanism apply support verb noun involve pair part speech describe special improvement esg lexical system work jeopardy first value general use esg include jeopardy fifth special jeopardy match noun frame verb frame esg lexicon populate complete slot frame open class word sensesvfor noun adjective adverb verb special effort work watson encode slot frame de verbal noun bcelebration bcelebrate verb relate noun way properly correspond verb frame relate noun verb way help relation extraction match question answer relation plus argument express verb place correspond noun kind correspondence see lexical entry talk slot frame noun sense correspond first verb sense noun slot nsubj nobj ncomp correspond respectively verb slot subj obj comp option specifie nsubj take default option agent n. agent option fille bby[-pps

ba talk john n option nsubj nobj fille bof[-pps ambiguity bthe love god disambiguate example ba talk john bthe choice department chairperson possessive noun premodifier fill nsubj implicitly example m. c. mccord et al 3 5 np bjohn talk mary book correspond vp bjohn talk mary book np bjohn fills nsubj noun btalk noun verb like frame morphologically derive verb bdiscussion[/bdiscuss btalk use word word etymologically related involve regular augmentation wordnet mention wordnet augment esg base lexicon main augmentation process completely automatic run

esg user recent version wordnet princeton augmentation aspect increase number entry mark augment lexicon systematically semantic type new entry consist proper noun certain multiword common noun wordnet vocabulary esg base lexicon extra semantic type marking follow esg attach mapping wordnet sense synset esg semantic type give lexical index word w sg sense frame w augmentation algorithm look wordnet sense w pos sense frame s go

wordnet hypernym chain s possibly find synset map esg type t mark t give sense frame w. process take account synset frequency number wordnet sense word avoid mark rare sense esg run augmentation augmentation increase parse accuracy score describe bevaluation augmentation base lexicon include addition kind relationship open class word nform vform ernform ervform verb v give compound feature nform n1 nm noun n1 nm nominal form v. example verb bdefer give feature nform deferral deferment deference feature vform provide inverse verb v give feature ernform n1 nm noun n1 nm agentive noun action v. example bcelebrate give ernform celebrator celebrant feature ervform provide inverse auxiliary lexical file augmentation come wordnet link derivationally relate word plus bit editing relation parsing provide esg parse output useful downstream watson match question possible answer concept express verb noun sg lexical coverage increase devicevchunk lexiconsvwhich use storage lookup scheme allow large number multiword entry jeopardy application esg use chunk lexicon ch.lx proper noun derive wikipedia anchor text approximately 1.4 million entry parse tree chunk entry find treat like multiword parse show underlie syntactic structure instance bse song sixpence ch.lx esg parse bshe sang sing song sixpence show figure 2 note bse song sixpence show multiword proper noun underlie vp structure show slot chsl chunk lexicon create unified medical language system national institutes health u.s. national library medicine http://www.nlm.nih.gov/

lat reward feature improvement specific jeopardy application augmentation esg base lexicon include illustration chunk lexical analysis underlie structure chunk m. c. mccord et al addition compound feature help parser choose np analysis lead question analysis well identification lexical answer type lat 1 lat term question indicate type entity ask lat reward feature noun sense n form latrwd r r floating point number syntactic analysis form np n head bthis bthese determiner certain constraint satisfie parse score np well c r c certain constant experimentally determine parse testing)vthus reward n typical lat configuration give n reward number r compute frequency table know lat occur jeopardy question r take sqrtðf = f0þ f frequency n

table f0 high frequency occur table latrwd add n occur table sg syntactic analysis tokenization segmentation morpholexical analysis main step syntactic analysis begin first step convert morpholexical analysis token word phrase phrase main data structure syntactic analysis work form bstarter phrase syntactic analysis step multiword agglomeration convert sequence word phrase phrase span multiword common multiword agglomeration base multiword entry lexicon basis general rule rule human name date literal number operation lookup chunk lexicon describe previous subsection occur stage multiword agglomeration nondeterministic sense syntactic analysis see agglomerate nonagglomerated span local syntactic analysis perform stage chunk multiword main sg syntactic analysis performedvvia left right chart parsing unit work datum type phrase main kind step binary modifier phrase fill slot adjacent matrix phrase left right slot try available complement slot list asl matrix adjunct slot associate pos matrix asl word phrase complement slot frame slot asl remove asl augment version matrix filler modifier attach asl augment matrix gain member slot extrapose modifier phrase e.g. handle bwh phrase coordination slot extrapose conjunct coalesce bjohn see mary hear train bthe train simultaneously fill obj slot conjunct modification process slot matrix try nondeterministically

fill obj slot conjunct modification process slot matrix try nondeterministically slot option try normally way slot-filler rule grammar rule base determine constraint fille slot plus option slot frame order actual filler phrase appear order sentence use slot option filler come remote part sentence bwhat see bwhat fill obj bsee remotely sg parsing use numerical scoring system intermediate final phrase assign parse score purpose final parse segment numerically rank application sg include watson use highest rank parse segment second parsing intermediate phrase analysis prune away parse space chart score fall low compare compete analysis span headword pos parse space pruning greatly increase efficiency time space affect parse outcome order parse risk overload long sentence turn pruning flag setting base parse score come source example general rule shell preference complement slot adjunct slot specific

scoring grammar reward penalty give lexicon latrwd feature addition parse scoring input bootstrappe way datum gather parse esg end chart parsing segment phrase build span segment sg form piece parse basis preference maximally span piece heuristic involve main step parsing sg analyze multiword unit way basis partial capitalization possible quoting sg form quote node proper noun node internal syntactic structure underneath instance sentence bhe appear meet press esg analyze bmeet press proper noun unitva quote nodevbut show vp structure underneath parse tree show figure 3 quote node formation similar result use chunk lexicon describe rely multiword lexicon quote node form regular parsing decision fit good overall parse special adaptation esg parsing process jeopardy question domain accommodate relatively frequent occurrence compare general english segment jeopardy np instead complete sentence example bnumber poem m. c. mccord et al 3 7

emily dickinson give permission publish lifetime idea np uniquely characterize answer heuristic boost score np analysis reward segment contain breference word bthis bhe bshe etc typically signal subphrase focus sentence question heuristic control flag default turn pas builder provide simplification abstraction esg parse remove detail design semantic distinction remove pas one subtle essential coarse grained distinction watson component use parse distinction component flexible require knowledge example passive active form assertion result pas esg parse active form logical argument passive verb logical subject bby[-pp pas building replace pp object preposition remove auxiliary bbe verb order produce exactly active form relation detector use pattern pas develop manually statistically active sentence apply passive sentence change contrast pattern write exactly parse different variant need handle active passive sentence illustrate example section pas builder process original text merely transform output esg simple comprehensive form pas segment label direct graph node citation form begin end character offset pos link esg parse node derive addition pas node derive esg parse node determiner additional string label indicate determiner determiner generally appear distinct node pas arc pas node node b label sg slot b fill a. expect meaningful

distinction lose move esg output pas lose meaning tend subtle e.g. difference emphasis difficult use effectively omit distinction easy write automatically induce pattern extract semantic question passage detect passage provide answer question example consider following hear edison invent phonograph 1877 hear edison invent phonograph 1877 hear edison invent phonograph 1877

hear edison invent phonograph hear phonograph invent edison sentence different meaning difference reflecte esg parse difference first particularly important semantically difficult use distinction general flexible way pas output sentence identical structure relation detection rule write pas match sentence match question align pas sentence e.g. bwho invent phonograph equally advantage pas disadvantage cause system draw incorrect conclusion e.g. conclude invent phonograph person merely invent advanced additional set text snippet treat identical pas odysseus bold clever bold clever odysseus parse show quote node m. c. mccord et al bold clever odysseus bold clever odysseus bold clever odysseus esg different dimension structure deep structure surface structure pas collapse dimension structure pas generally follow deep structure type slot encode surface structure pas node lemma form feature roughly correspond lemma esg parse esg special encoding derivational morphology circumstance e.g. bnon mormon text encode bnon+mormon esg lemma pas remove variety node esg omit pas link node parse tree instead redirect related node pas example link auxiliary verb instead treat link correspond main verb link omit relate node completely omit example link auxiliary verb corresponding main verb omit pas specific type node omit follow auxiliary verb include auxiliary verb indicate closed class node introduce vp infinitive bto marker bthat introduce determiners one special whitelist bhigh semantic determiner e.g. possessive pronoun determiner indicate negation note string value determiner retain string value pas node attach form bbe predicate form bbe predicate adjective addition pas provide simplified pos taxonomy esg pos completely omit note esg subconj subordinate conjunction mark prep preposition pas esg qual qualifier mark adv adverb pas simple taxonomy sufficient key coarse

distinction allow downstream processing distinguish noun verb kind modify connect word make subtle distinction word class esg provide esg coordinate conjunction e.g. band mark pos element coordinate separate cord feature indicate conjunction pas pos node cord

coordinate separate cord feature indicate conjunction pas pos node cord pos constituent mark constituent change convenient component perform local processing specific node example component want iterate noun sentence want treat conjunction noun simply check pos node have separate check node coordination summary pas chandelier question esg parse show figure 1 look 2 subj:1 comp:3

4 lconj:2 rconj:9 predicate use 9 subj:1 obj:11 vadv:5 vadv:7 item 11 nrel:17 determiner 12 objprep:13 15 ndet:14 determiner derive 17 obj:15 comp:12 notation provide i d number node first number open parenthesis follow zero label argument consist label colon id number target node example bchandelier id 1 buse id 9 first label argument indicate subject id 1 i.e. bchandelier subject buse id number merely artifact present pas paper deepqa pas encode uima cas data structure label pointer node id number need key difference pas example corresponding esg parse figure 1 include follow esg parse figure 1 show surface structure line deep structure logical argument pas distinction structure show label logical argument example node bderive 17 pas argument derive deep logical structure correspond node esg parse contrast node bbut 4 argument pas label blconj brconj derive esg surface structure node present parse omit pas specifically following node parse omit determiner bthese auxiliary verb node bdo bbe determiner btheir possessive pronoun drop pas builder consider high semantic determiner m. c. mccord et al 3 9 note early esg parse argument frame include entry label u bunfilled bunknown filler e.g. derive(17,u,15,12

indicate bderive unfilled slot subject pas explicitly encode information noun determiner attach parse determiner string label pas esg parse set feature show rightmost column figure 1

encode pas expect pas user want feature examine correspond node esg parse pattern base relation extraction section analysis provide deep parsing facilitate pattern base relation extraction relation extraction identifie domain specific semantic level relation sentence argument typically type entity example semantic relation authorof actorin bornon refer deep relation abstract away esg pas analysis abstraction process answer lookup 2 passage scoring 3 structure inference 14 great variability lexical syntactic relation express instance sample jeopardy question instance authorof relation list clearly anticipate write individual pattern productive way recognize relation representational device esg pas analysis turn example handle single rule rule require 1 1936 write play bthe boy david actress play title role 2 bear winsted practice law connecticut write bunsafe speed 3 bfrench connection actor coauthore 1999 novel bwake perdido star 4 walter mosley pen mystery detective easy rawlins search woman 5 december 1513 write francesco vettori bcompose little work fon princedom 6 bmanly 19th century realist pen work like badam bede bfelix holt bdaniel deronda 7 robert louis stevenson fall love fanny osbourne married woman later write tale son 8) friend refuse destroy kafka work write historical novel tycho brahe 9 live vermont kipling begin write tale orphan son irish soldier india 10 baccorde source first game book write bshort treatise game whist 11 bthe lair white worm bdracula write dubliner 12 bsomnium early work science fiction write german publish posthumously 13 rip tale appear bthe sketch book write 14 originally write alexander pushkin poem russian novel later turn opera 15 author bjazz btar baby first black american win nobel prize literature 16 author bfathers sons first russian widely read admire europe 17

beckett major dramatist btheater portray bewildered anxious 18 edmund white write definitive biography french thief novelist playwright bthe maids example 1 2 illustrate relatively straightforward authorof expression verb bwrite immediately precede subject potentially author follow object refer title allow lexical variability illustrate observation suggest pattern follow authorof author writeverb work issue recognize component pattern text linear

sequence base framework problem identify headword constituent complex phrase orthogonal framework need sensitive alternative lexical cue authorship example 3 6 cf bcoauthor bpen bcompose bnovel bmystery bbestseller etc indicator authorof additional requirement manage identify specific element trigger pattern match e.g. bpenned 6 distinct work argument signal coordination brobert louis stevenson author argument bwrote 7 difficult detect give detract effect intervene essentially problem linear pattern word blook alike contrast esg pas provide analysis backbone variability textual expression reduce manageable number patternsvbecause target deep constituent surface word token example syntactic analysis question identifie exactly element typically subj obj need examine express possible authorof relationship m. c. mccord et al numerous aspect esg syntactic parse play structured rendering pattern instance consider following multiword name entity recognize single node make natural target match cf brobert louis stevenson 7 bwalter mosley 4 b complex np relational noun internal structure assignment help semantic interpretation cf bnovel fwake perdido star bnovel head noun material easily associate title 3 c quote material properly handle large context get structure analysis suitably integrate overall parsevwhich allow pattern apply bacross quotation mark h bcompose little work fon princedoms_[i 5 d analysis coordinated construction allow matching framework develop uniform analysis node relation instance rule need separately cater single conjoin relation argument rule emit distinct work argument bpenned 6

supply right author argument bwrote 7 cord association bfell love esg go systematically identify explicitly label long distance relationship constituent bring diverse text form close instance arbitrary text interpose head noun covering verb see need connect brobert louis stevenson bwrote 7 example 8) similarly require identify- e bthis friend potential author argument distance apart writeverb general esg attempt share argumentsvwhere semantically identicalvacross multiple clause boundary syntactic context differ variety way bkipling subject bbegan bwriting 9 bbook object bwrote bwas 10 clearly value pattern outline early

interpretation structuralvas oppose linearvtemplate assume notation nodea > dependencylabel > nodeb refer dependency link parse node elaboration early pattern target dependency tree single pattern conjoin label link anchor verb node match example far authorof writeverb > subj > author writeverb > obj > work choose define relation detection pattern pas representation pas builder preserve information argument cluster node introduce normalization example additional normalization identical representation pas active passive form phrase esg retain bby marking rule need duplicate target dependency tree contrast apply pas rule unmodified detect relation variety passive construct simple e.g. conjunction 11 appositive 12 relative clause 13 finite vp premodifier postmodifier 14 forth additional simplification pas esg contribute clarity rule writing keep number rule include mark coordinated construction unique cord pos suppress distinction modifier left right head packaging modifier label synchronous list section bpredicate argument structure illustration point consider expand current rule base pattern detect instance authorof relation express mean relational noun example 15 18

key bof[-pp attach author predicate syntactically analyze nobj nprep construction potentially lead separate rule esg consequently pas dependency label pas convention uniform access node modifier modifier pair label attach head allow single rule loosely state conjunction dependency link share bof node authorof author > label > of"-arg of"-arg > objprep > work label act wildcard nobj nprep rule identify argument list argument node author govern bof preposition link np head pass work discussion show convention analytic style expose esg pas facilitate abstraction relatively small set rule whichvby target constituent argument dependency link labelsvare capable handle variability relation expression text use watson note introduction paper esg parse pas semantic relation different component watson consider example clue early text bchandelier look great nowadays usually use item derive section describe way parse information derive parse m. c. mccord et al 3 11 question analysis 1 attempt identify question focus term clue substitute correct answer produce true statement

analysis apply variety rule parse assessment example clue focus bthese item infer fact word bthese determiner question analysis attempt identify lat head focus word closely connect focus parse example clue lat bitems head focus complex clue lat disjoint passage score component 3 use result derive esg parse specifically pas semantic relation question text text passage believe contain support evidence candidate answer component use information determine strength evidence passage assessment involve determine candidate answer passage align focus question example pas show focus bthese item object verb buse subject bchandelier modifier bnot passage contain verb buse object candidate answer subject bchandelier modifier bnot passage scorer assign high score candidate answer passage similar word e.g. synonym similar structure passage scorer provide partial credit different passage scorer different restriction term connection term e.g. pas semantic relation match receive different level credit 3

prismatic knowledge base 11 include statistic relate set word connect specific pas structure semantic relation example prismatic record word bchandelier subject verb buse large corpus give word object verb buse subject bchandelier data finde 2 evaluate 3 10 component 10 apply esg pas annotator analyze content source answer type information source typing information wikipedia category short text string describe type entity page example wikipedia page btoronto ohio category bpopulated place jefferson county ohio give label want conclude btoronto ohio legitimate answer question ask bplace want conclude btoronto ohio legitimate answer question ask bcounty word appear category text esg parse show bplaces syntactic head np bpopulated place jefferson county ohio consequently type coercion algorithm conclude category assert toronto ohio place conclude toronto ohio county note early small sample watson use esg parse information derive information word connect question variety evidence source vital operation deepqa technology pervasively watson esg evaluate parse correctness test set 100 jeopardy question segment b 100 segment wikipedia test set randomly choose esg training esg test ne name entity chunk lexicon extract wikipedia

mention evaluate early version esg esgbase test set esgbase version date prior commencement work watson charniak parser cp reranke version mark johnson 15 evaluate test set cp deliver different kind parse tree penn treebank ptb style tree view dependency tree straightforward wayvfor example

replace s np vp v vp np v node n dependency parse tree call correct mother node exhibit n correct pos n correct pos take level ptb pos score parser percentage correct node deliver test set automatically rank parse segment recall 100 parser deliver node datum node test set score give table 1 evaluation require cp show ptb cp weak show coordination structure particularly np beginning head noun bnp np usually show flat way m. c. mccord et al cp coordination surface structure esg show penalize get wrong flat cp np front coordination penalizedvexcept situation hold possible choice conjunct cp np analysis weak include postmodifier np end flat np list bthe brothers grimm difficult tell head noun cp weak inconsistent show capitalize noun proper common noun show adjective structure productive hyphenation note node correctness matter surface structure scoring table 1 reward esg extra information provide deep structure slot-filler marking rich semantic type marking etc additional information useful watson score test set late version esg wordnet base augmentation describe result 91 outstandingly low 92 wordnet augmentation measure value considerably great semantic type marking useful downstream component improvement esg parsing efficient process 5,000 word second standard laptop particularly valuable jeopardy application efficiency owe careful implementation c care write lexicon syntax rule parse space prune algorithm test 6,712 jeopardy question segment esg run total 17 second test set machine linux xseries machine cp run 1,763 second thread

104 time slow esg cp run thread thread run 57 time slow esg machine core three- threaded run cp take long threaded cp run esg equip run thread sentence parse process large corpus easy divide corpus piece run different piece different core esg small footprint binary code

datum count chunk lexicon occupy 5.7 mb work memory normally 52 mb version esg watson easily compile run smartphone formally evaluate pas builder give pas builder provide simplification esg parse believe correctness result esg provide adequate sense accurate pas exact percentage correct node slightly different pas few node give node drop general pas correct extent esg parse correct believe conduct distinct formal evaluation pas relation extraction formally evaluate 9 esg rule base linguistically orient head drive frame orient largely lexicalist probably closely relate head drive phrase structure grammar hpsg 16 lexical functional grammar lfg 17 commonly computational grammatical system sg first computational head drive lexicalist grammatical system first work 1978 publish 1980 4 combination sg surface structure deep structure analogous lfg constituent structure functional structure difference sg parse tree surface deep structure combine tree natural sg function slot play crucial role parsing useful aspect analysis combine element surface structure play role logical form happen e.g. generalized quantifier focus adverb pas usefully continue combination commonly statistical parser charniak parser 15 collins parser 18 stanford parser 19 20 minipar 21 sg closely relate dependency parser minipar choose compare esg charniak parser relatively wide use development regimen esg regular testing observation problem make fixe general possible introduction new construction appropriate constant regression testing set

20,000 segment source set overlap test set b bevaluation section use tree bank imitate tree bank stifle introduction useful new construction exist tree bank term bpredicate argument structure refer wide variety parse like construct usually focus semantic aspect analysis syntactic one example pas verbnet 22 use thematic role agent patient assign thematic role system need sort lexical resource describe role apply specific word specific syntactic construct e.g. 22 and/or label training datum illustrate mapping example e.g. 23 pas instantiation simply use syntactic label directly need perform m. c. mccord et al 3 13 mappingve.g 24 pas describe paper note pas include esg complement slot-filler argument semisemantic nature correspond

thematic role argument pas instantiation sort necessarily require extensive lexical knowledge simple implement obviously provide extensive semantic insight provide semantic orient level analysis parse usually remove parse ingredient important syntactically deepqa project include component

determine thematic role relate term text consider distinct level analysis specifically bshallow semantic relation detection 3 separate pas identification thematic role distinct component provide separate complementary level analysis downstream component use separately conclusion future work esg produce informative parse tree exhibit deep structure surface structure unified way esg broad coverage good efficiency esg parse particularly deep structure feed simplified abstract analysis pas result esg pas builder downstream component watson example pattern base relation recognition use syntactic analysis generate esg simplified abstract pas builder provide illustration give overview esg pas analysis indicate central role slot slot frame emphasize new ingredient develop work watson special jeopardy task general value future work sg include follow 1 expansion semantic type system use parsing 2 incorporation word sense disambiguation probably sense granularity wordnet 3 indication parse tree scoping generalize quantifier focus adverb etc 4 development specialized lexicon method handle large lexicon 5 continued improvement coverage sg regression testing business machines corporation united states country productions inc. trustees princeton university wikimedia foundation linus torvalds united states country 1 a. lally j. m. prager m. c. mccord b. k. boguraev s. patwardhan j. fan p. fodor j. chu carroll bquestion 2 j. chu carroll j. fan b. k. boguraev d. carmel d. sheinwald c. welty bfinding needle haystack search 3 j. w. murdock j. fan a. lally h. shima b. k. boguraev 5 m. c. mccord bus slot modifier logic grammar 6 m. c. mccord bheuristics broad coverage natural language parsing proc arpa hum lang technol workshop 1993 9 c. wang a. kalyanpur j. fan b. k. boguraev d. c. gondek brelation extraction scoring deepqa 10 j. w. murdock a. kalyanpur c. welty j. fan d. a. ferrucci d. c. gondek l. zhang h. kanayama btyping candidate

11 j. fan a. kalyanpur d. c. gondek d. a. ferrucci 12 g. a. miller bwordnet lexical database english 13 c. fellbaum wordnet electronic lexical database cambridge ma mit press 1998 14 a. kalyanpur b. k. boguraev s. patwardhan j. w. murdock a. lally c. welty j. m. prager b. coppola a. fokoue nkoutche l. zhang y. pan z. m. qiu 15 e. charniak m. johnson bcoarse to-fine n best parsing maxent discriminative reranking proc 43rd annu 16 c. pollard i. a. sag head drive phrase structure chicago il univ chicago press 1994 17 r. m. kaplan j. bresnan blexical functional grammar formal system grammatical representation mental representation grammatical relations j. bresnan ed 18 m. collins bhead drive statistical model natural language 19 d. klein c. d. manning baccurate unlexicalized parsing proc 41st meeting assoc comput linguist 2003 20 m.-c. de marneffe b. maccartney c. d. manning bgenerate type dependency parse phrase structure 21 d. lin bdependency base evaluation minipar proc 1st int conf lang resources eval vworkshop evaluation parsing systems granada spain 1998 22 k. k. schuler bverbnet broad coverage comprehensive verb lexicon ph.d. dissertation univ pennsylvania philadelphia pa jan. 2005 paper aai3179808 m. c. mccord et al 23 k. kipper m. palmer o. rambow bextending propbank verbnet semantic predicate proc amta workshop appl interlinguas tiburon oct. 2002 online available http://www.mendeley.com/research/ 24 r. krestel r. witte

s. bergler bpredicate argument extractor pax proc lrecvnew challenges receive july 27 2011 accept publication november 17 2011 michael c. mccord watson research center yorktown heights ny 10598 usa semantic analysis integration department t. j. watson research center receive ph.d. degree mathematic yale university new haven ct spend year institute advanced study princeton nj initial research mathematic topology foundation move computer science natural language processing emphasis originate slot grammar parsing system apply machine translation grammar checking watson question answer system author coauthor 47 referee article book j. william murdock watson research center yorktown heights ny 10598 usa deepqa

research team t. j. watson research center 2001 receive ph.d. degree computer science georgia tech atlanta member ashok goel design intelligence laboratory work post doctorate david aha u.s. naval research laboratory washington dc research interest include natural language semantic analogical reasoning knowledge base planning machine learning self aware artificial intelligence branimir k. boguraev watson research center yorktown heights ny 10598 usa semantic analysis integration department thomas j. watson research center receive engineering degree electronics higher institute mechanical electrical engineering sofia bulgaria 1974 diploma ph.d. degree computer science 1976 computational linguistic 1980 respectively university cambridge cambridge u.k. work number u.k./e.u. research project infrastructural support 1988 work resource rich text analysis 1993 1997 manage natural language program apple advanced engineering large scale business content analysis recently work jeopardy challenge algorithms team develop technology advanced question answering dr. boguraev author coauthor 120 technical paper 15 patent recently executive editor cambridge university press book series studies natural language processing member editorial board computational linguistics journal semantics continue serve found editor journal natural language engineering member association m. c. mccord et al 3 15