

Jude: la realizazion di un program di traduzion automatiche dal talian al furlan

Sandri Carrozzo, Franz Feregot, Dree Mistrut
SERLING soc. coop.
publicât in Ce fastu? 2010-2, p. 269-280

Dal 2003 la cooperative Serling e lavore su sistemis di traduzion automatiche dal talian al furlan, nomenâts Jude, par vie che a son nassûts par “judâ” il tradutôr tal so lavôr e cun riferiment a Jude, l’apuestul traditôr, viodût che al è ben cognossût il zûc di peraulis che al fâs di ogni tradutôr un traditôr e che al vûl che, voltant cualsisei test di une lenghe a di une altre, si fasi ancje un tradiment rispjet al originâl. In chest articul si presente une sintesi des carateristichis di chescj sistemis e dai risultâts otignûts.

1.1 La traduzion automatiche: definizion e finalitâts

Par traduzion automatiche si intint la traduzion di une lenghe a di une altre fate di un program informatic. In gjenerâl la ativitât de traduzion e je une vore complesse e in tradutologie si calcole che nissune traduzion di un test avonde complès, nancje chê fate di une persone cun competencis professionâls une vore altis, e sedi perfete e definitive.

In cheste suaze si pues capî che l’obietîf pratic de traduzion automatiche, che si puedi pardabon realizâ cui mieçs e cui cognossiments teorics e tecnicos dal di di vuê, nol è chel di cjapâ il puest dai tradutôrs in cjar e vues, fasint traduzions za dal dut coretis, ma ben chel di judâ i professioniscj de traduzion a lavorâ in maniere plui svelte e, in cualchi câs, plui rigorose.

Di là di chest obietîf, che di fat al è avonde dificil, a son programs informatics, che in gjenerâl si puedin doprâ a gratis in Internet, che si contentin di dâ al utent une traduzion une vore arossimative, dispès permetint ae persone di induvinâ, plui che no di cognossi cun avonde sigurece, il contignût di tescj scrits intune lenghe foreste che no cognôs o che no cognôs avonde ben.

Par ce che al tocje in particolâr il progjet Jude la finalitât no je chê di rivâ ae perfezion, che cemût che o vin dite no je ancjemò un obietîf realistic, ma nancje chê di jessi un program une vore arossimatîf e dome dimostratif: Jude al vûl jessi un strument util pe traduzion a nivel professionâl.

1.2.1 Diviers sistemis di traduzion automatiche

Par rivâ al risultât che un program informatic al puedi fâ une traduzion di une lenghe a di une altre si pues sielzi tra diviersis strategiis, che a rispuindin a dibisugnis diferentis e che si fondin su basis diferentis: concezions linguistichis, nivei tecnologjics, grât di difference tra lenghis, risorsis di timp, di competencis e di bêçs a disposizion e v.i. a puartin a frontâ la traduzion automatiche in manieris une vore diviersis.

Tra i diviers modei di traduzion automatiche chei plui doprâts a puedin jentrâ in chestis categoriis:

- traduzion direte
- traduzion cun corpus:
 - traduzion statistiche
 - traduzion daûr di esemplis
- traduzion daûr di regulis
 - trasferiment
 - interlingua

1.2.2 La traduzion direte

La traduzion direte e je il sisteme plui sempliç di traduzion automatiche: in gjenerâl e funzione metint in vore une semplice liste di corrispondencis di peraulis. La sô precision e je une vore limitade, ma e à il vantaç che realizâ programs di cheste sorte al è avonde sempliç e svelt. In cualchi câs si fevele di traduzion direte ancje pai sistemis di traduzion a trasferiment plui rudimentâi. Cheste strategjie e je stade la prime a jessi cjapade in considerazion pe realizazion di Jude.

1.2.3 La traduzion statistiche

La traduzion automatiche statistiche si fonde su la relazion tra corpus paralêi in dôs o plui lenghis: il tradutôr automatic, devant di un test gnûf di voltâ, al controlarà tai siei corpus se al à za lis secuencis di peraulis o, se di no, lis singulis peraulis che a comparissin tal test di voltâ e al metarà in vore chês stessis soluzions che al cjate cun plui frecuece tai siei corpus. I vantaçs di chest sisteme a son che il program di realizâ nol è masse complès e duncje i coscj a son limitâts, ma par vê un funzionament suficient si à bisugne di corpus paralêi di milions e milions di peraulis: in chest moment tra la lenghe taliane e furlane nol è pussibil dâ dongje corpus di chestis dimensions, duncje cheste soluzion no somee che si puedi aplicâle cun bon sucès ae traduzion automatiche dal talian al furlan.

1.2.4 La traduzion daûr di esemplis

Ancje la traduzion automatiche daûr di esemplis si fonde su corpus paralêi, ma invezit di riprodusi in maniere mecaniche ce che si cjate za in memorie e cîr di aplicâ regulis di analogjie. Chest i permetarès di funzionâ cun corpus paralêi simpri grancj, ma di dimensions plui piçulis rispiet ai tradutôrs statistics. Sistemis di cheste sorte a son doprâts in programs une vore cognossûts di traduzion assistude, che si diferenziin de traduzion automatiche parcè che a domandin l'intervent dal tradutôr uman ancje intant dal procès di elaborazion e no dome in fase di revision, o in programs di traduzion automatic tra lenghis avonde differentis tra di lôr: par esempli sistemis di cheste sorte a son stâts valutâts chei plui adats par disvilupâ tradutôrs automatics tra basc e castilian e basc e inglès. Viodût che la plui part dai tradutôrs automatics di cheste sorte a jerin ancjemò in fase sperimentâl e che in ogni câs si fondin su corpus di

dimensions plui grandis di chei che disponibii tra talian e furlan (tal cās di basc e inglēs si veve tacāt cuntun corpus di 270.000 frasis paralelis), cheste linie strategjiche no je stade cjapade in considerazion tal moment di programā Jude.

1.2.5 La traduzion a trasferiment

A son diviers sistemis di traduzion automatiche a trasferiment, ma in gjenerāl si pues di che, invece di fondāsi suntune grande base di dāts di corpus paralel, si fondisi su regulis che a permetin di analizā di un pont di viste morfologjic e in cualchi cās ançe sintatic o fintremai semantic il test di partence, si cīr di ricognossi ecuivalencis cu la lenghe di traduzion midiant di un vocabolari e di regulis gramaticāls e duncje di creā la traduzion.

In gjenerāl si fevele di trasferiment superficial se si fermisi ae analisi morfosintatiche, di trasferiment profund se si da une vore di impuartance ae analisi semantiche. Tal cās di traduzion a trasferiment une vore rudimentāl si fevele ançe di “traduzion direte”.

Rispiet ae traduzion che si fonde su corpus, chē a trasferiment e à il svantaç di domandā la formulazion di regulis complessis e la compilazion di un vocabolari gnūf par ogni cubie di lenghis e di domandā un sfuarç linguistic e informatic plui alt e specific. In compens il trasferiment superficial al è il sisteme che fin cumò al à dimostrāt di dā risultāts miōr inte traduzion di lenghis no masse lontanis tra di lōr; il trasferiment profund al ven doprāt invece tra lenghis di famee diferente. Cheste strategjie, e te ultime version in particolar il trasferiment superficial, e je chē che e je stade sielzude tal program Jude.

1.2.6 La traduzion cun interlingua

Ançe la traduzion cun interlingua, o cun lenghe intermedie, si fonde su la analisi morfologjiche, sintatiche e dal origjināl, ma invece di passā al trasferiment te lenghe di traduzion si fās une raprezentazion intune lenghe intermedie artificiāl che e vedi dutis lis pussibilitāts morfologjichis, sintaticis e semanticis sedi de lenghe di partence che de lenghe di traduzion. Chest sisteme, plui complēs e impegnatif di chel di trasferiment par vie de “invenzion” de lenghe intermedie, al permet di vē risultāts cun mancūl sfuarçs tal cās di tradutōrs automatics che a voltin tal stes moment in plui lenghis e in plui direzions e al permet di tratā lenghis lontanis tra di lōr par struture.

Viodūt che la lenghe taliane e chē furlane a son dutis dōs de famee neolatine e che in chest moment nol è un fuart interès comerciāl tal voltā di plui lenghis al furlan e dal furlan a plui lenghis, cheste strategjie no je stade doprade tal cās dal program Jude.

1.2.7 Tradutōrs cun sistemis ibrits

Viodūt che ducj i sistemis di traduzion automatiche presentāts fin ca a àn aspiets positifs, ma ançe difiets o svantaçs impuartants, la tindince che e je daūr a disvilupāsi e je chē di rivā a programs cun funzionament ibrit, vāl a dī che a doprin in maniere coordinade plui di une des strategjiis presentadis fin cumò. Al è pussibil che intal futūr ançe il program Jude si indreci par cheste strade.

2. La elaborazion dal tradutôr automatic talian-furlan Jude

Il tradutôr automatic talian-furlan Jude al è stât elaborât tacant dal 2003 de Serling, une cooperative di servizis linguistics nassude tal 2001 e che e je stade sburtade a elaborâ chest strument di un interès comerciâl: di fat lis traduzioni dal talian al furlan a àn simpri vût un pês une vore impuartant tal complès de sô ativitât e sveltî e miorâ il procès di traduzion al è diventât in curt une prioritât.

2.1 Jude 1: traduzion direte, cun cualchi justament e interazion dal tradutôr uman

Il prin progjet di doprâ i struments informatics par sveltî i lavôrs di traduzion dal talian al furlan al è stât metût in vore intal 2003, par rispuindi a une incesitate improvise de ativitât di traduzion, prin di dut par tradusi tescj ripetitifs e cuntun lengaç cristalizât cemût che al è chel di leçs e regolaments.

Jude 1 al jere un program di traduzion direte, avonde sempliç e par di risorsis linguistichis: di fat al è rivât a vê un vocabolari di mancûl di 10.000 peraulis. No si veve a disposizion regulis gramaticâls pe analisi e nancje un vocabolari masse grant, lis peraulis de lenghe di partence a jerin metudis intune liste di corispondence cun peraulis furlanis, strutturade a siet nivei. Tal câs di formis ambiguis, intun ciert numar di câs a jerin procediments che a puartavin a produci une soluzion corete midiant di passaçs sucessîfs (par un massim di 7), ma te plui part dai câs di ambiguitât, se il program nol cjatave riferiments utii, al domandave al tradutôr uman di sielzi tra lis soluzions pussibilis.

Ancje se il materiâl linguistic computazionâl di Jude 1 al jere une vore limitât, si pues di che, metint in relacion lis oris impegnadis par realizâlu e chês che al à fat sparagnâ tal prin an di funzionament, al è stât un sucès e che tal 2003-2004 nol jere pussibil rivâ cun facilitât a struments plui perfezionâts. Pal so funzionament, util ma limitât, Jude 1 al è stât doprât de Serling tant che strument di traduzion a ûs interni, ma nol è stât presentât al public.

2.2 Jude 2: traduzion direte di peraulis etichetadis e cun regulis

Cul passâ dal timp si è viodût che, ancje se Jude al funzionave avonde ben, al veve limits di impostazion che si veve di cirî di miorâ: invistî dome su la cressite cuantitative dal numar di peraulis de base di dâts nol jere convenient. Il limit principâl di un pont di viste pratic al jere che il program al domandave une vore dispès ae persone cemût compuartâsi tra plui alternativis: par esempi inte traduzion di dutis lis formis verbâls di III persone singolâr si domandave se zontâ il clitic *al* o *e* daûr che il subiet al fos masculin o feminin.

Par sveltî la procedure si à cambiât il program di plante fûr, elaborant di fat un strument informatic gnûf, Jude 2, che al previodeve la etichetadure di dutis lis peraulis talianis e di dutis lis peraulis furlanis de base di dâts. In struc si tratave ancjemò di traduzion direte, ma si deve al program i riferiments par analizâ il contest dulà che si cjatave ogni peraule e di risolvi in autonomie un numar une vore plui grant di ambiguitâts.

Ancje Jude 2 al traduseve passant par 7 nivei progressifs, ma al veve une struture che e previodeve par ogni peraule la pussibilitât di zontâ regulis che a indreçassin la traduzion daûr des peraulis che si cjatavin prime o dopo di chê tratade. Par esempi la peraule *è* e jere voltade cun *al* è se prime e veve un sostantif o un adietif o un pronon masculin singolâr, o ben se dopo e veve un participi o un adietif masculin singolâr; invezit e jere voltade cun *e je* se chês stessis parts dal discors a jerin femininis singolârs. Tal câs di ambiguitâts impossibilis di risolti daûr des regulis inseridis Jude 2 al domandave al tradutôr uman di sielzi tra plui pussibilitâts.

Jude 2 al è rivât a vê une base di dâts cun plui o mancul 20.000 peraulis etichetadis: il so funzionament al jere une vore superiôr a chel di Jude 1.

Il so limit principâl al jere chel di fondâsi suntune base di dâts cun corrispondencis tra peraulis etichetadis e no tra lemis etichetâts: in ceste maniere l'inseriment di gnovis formis e duncje il slargjament dal numar di peraulis che al podeve voltâ al diventave lunc e, a di un ciert pont, pôc economic. Ancje Jude 2 al è stât doprât dome tant che strument interni de Serling e nol è stât presentât al public.

2.3.1 Lis cundizions par realizâ un tradutôr a trasferiment

I limits di Jude 1 e Jude 2, plui che altri a derivavin des pocjis risorsis di materiâl linguistic util a disposizion de Serling: di fat par rivâ a un vêr tradutôr a trasferiment, in plui des competencis linguistichis e informatichis a coventavin basis di dâts di lemis etichetâts sedi de lenghe di partence, sedi de lenghe di traduzion, metudis in relacion di corrispondence tra di lôr.

La situazion par ce che al tocjave la part furlane, lenghe di traduzion te prospettive dal tradutôr automatic, e jere avonde buine: tal 2001 i stes autôrs di Jude 1 e Jude 2 a vevin elaborât par Informazion Furlane il Coretôr Ortografic Furlan, fasint la sielte di impuartance fundamentâl di fondâsi su regulis morfologjichis e no suntun sisteme statistic: duncje e esistev une base di dâts di lemis furlans etichetâts. Simpri chês stessis personis a vevin tornât a ordenâ e elaborâ chest materiâl par furnî al Grant Dizionari Bilengâl Talian-Furlan dal Centri Friûl Lenghe 2000 la funzion dai cuadris flessionâi.

Tal 2007 la Serling e à comprât di Informazion Furlane i dirits a doprâ a elaborâ lis basis di dâts dal Coretôr Ortografic Furlan e dal Dizionari Ortografic Furlan: cussì, cuntun altri svilup e cuntun completament dal incasement des flessions, ancje par ce che al tocjave lis formis che no si visualizin tal GDBTF, la part fundamentâl de lenghe di traduzion di doprâ intun tradutôr a trasferiment e jere pronte.

Par ce che al tocje la part de lenghe di partence, il talian, la Serling e à valutât la pussibilitât di comprâ lis basis di dâts necessariis di cjasis editoriis o centris di ricercje talians, ma dopo di une prove a campion si è viodût che cu la esperience fate tal tratament de lenghe furlane al jere plui convenient produsi la base di dâts taliane che no comprâle: scuasi in contemporanie cui lavôrs necessariis ae publicazion dal DOF di cjarte, e je stade fate la etichetadure dai lemis talians e la formulazion des regulis morfologjichis pe produzion di dutis lis formis fletudis: cussì tra la fin dal 2007 e il principi dal 2008 e tacave la elaborazion di Jude 3, che e je stade completade ae fin dal 2009.

2.3.2 Jude 3: traduzion a trasferiment superficial

Cun Jude 3 si è rivâts a un vêr tradutôr a trasferiments superficial, dotant in cheste maniere la lenghe furlane di un strument che, a nivel potenziâl, al è chel bon di dâ i risultâts miôr te traduzions de lenghe taliane.

La realizazion di Jude 3 e je lade daûr di modei aromai canonicis a nivel internazionâl te realizazion dai tradutôrs di cheste sorte, duncje il program al è componût di modui che a fasin chestis funzions:

- Tokenizazion: a vegnin isoladis lis componentis lessicâls dal test par che dopo a puedin jessi processadis une par une; cheste funzion e pues someâ semplice, par vie che lis peraulis a vegnin scritis dividudis di spazis o di simbui di interpunzion, ma a son ecezions che a rindin complesse ancje cheste operazion (par esempi peraulis scritis cun simbui che però no dividin dôs unitâts lessicâls: nord-est, socio-linguistico, www.sitointernet.it, e v.i.).
- Analisi morfologjiche: a ogni unitât lessicâl si associe un trinomi (o plui trinomis in câs di ambiguitât) formât di leme - categorie gramaticâl - determinazion flessionâl: par esempi a *case* al sarà associât il trinomi [casa_SF_FP] vâl a di: leme casa - categorie gramaticâl sostantîf feminin - determinazion flessionâl feminin plurâl; a *porto* a saran associâts i trinomis [porto_SM_MS; porgere_VTR_PPM1; portarsi_VRIF_IP1; portare_VTR_IP1], stant che al pues vignî di plui lemis, che a puedin jessi un sostantîf masculin o trê verps e che a puedin jessi di diversis determinazions flessionâls: masculin singolâr, participi passât masculin singolâr, indicatîf presint di prime persone. Par fâ cheste operazion Jude al dopre une liste di plui di 58.000 lemis, ognidun associât al so arbul flessionâl complet, e in cheste maniere al cognòs plui di 3.400.000 formis talianis.
- Disambiguazion: viodût che, come tal câs di *porto*, une vore di peraulis a àn plui pussibilitâts di associazion a lemis, categoriis gramaticâls e morfologjichis, e covente une disambiguazion che e lavore a trê nivei in chest ordin:

- Disambiguazion gramaticâl: si etichetin lis parts dal discors e si metin in pratiche regulis che a permetin dome ciertis secuencis, coretis, e invezit a eliminin chês altris cussì te secuencia *il porto*, dopo de analisi morfologjiche e de etichetadure des parts dal discors a saressin pussibilis lis secuencis:

ART_M_S NON_M_S

(articul masculin singolâr - non masculin singolâr)

ART_M_S V_TR_PPA_M_S

(articul masculin singolâr - participi passât di verp transitîf)

ART_M_S V_RIF_PPA_M_S

(articul masculin singolâr - participi passât di verp riflessîf)

ma daûr des regulis di chest modul di disambiguazion dome la prime secuencia e sarà acetade, viodût che al program i è stât insegnât che chês altris dôs secuencis no son ametudis de sintassi taliane. In cheste maniere il program al savarà che in cheste situazion la peraule *porto* e à di jessi associate a porto_SM_MS. Pe gjeston di chest modul a son necessariis une vore di

regulis, in part gjeneradis in maniere automatiche, in part compiladis a man. Pal moment Jude al disambigue a chest nivel metint in vore un miâr di regulis.

- Disambiguazion dal leme: a son câs di ambiguitât che si riferissin dome al leme, intant che la determinazion flessionâl no je influente. Par esempi la peraule *allevino* e sarà simpri un coniuntif presint di tierce persone plurâl, ma e podarà derivâ dal verp *allevare* o dal verp *alleviare*. In chescj câs Jude al apliche un modul di regulis di ricognossiment dal contest semantic: cussì tal câs che te stesse frase si cjati il leme *bestiame*, al deciderà che *allevino* al è une forme di *allevare*, invezit tal câs che al cjati un leme tant che *dolore*, al deciderà che al derive di *alleviare*. Câs di ambiguitât di cheste sorte si cjatin cun mancual frequence e pal moment Jude ju fronte aplicant mancual di 2000 regulis.
- Disambiguazion semantiche: a son câs di ambiguitât che a tocjin lemis omografs e che a son associâts a chês stessis categoriis gramaticâls e morfologjichis, ma che a àn une difference di significât fundamentâl, par esempi *pesca* al pues valê par *pescje* o ben par *pierçul*. Ancje in chescj câs Jude al dopre un modul di regulis che si fondin sul contest, cussì al deciderà che se plui o mancual dongje di *pesca* al cjate lemis tant che *salmone*, *trota*, *mare*, *fiume* e v.i., al assegnarà a *pesca* il prin significât, invezit se al cjatarà lemis tant che *macedonia*, *frutta*, *succo*, *albicocca* e v.i. al assegnarà a *pesca* il secont significât. Di ambiguitâts dal gjenar in realtât no si 'nd cjate masse dispès, pal moment Jude lis fronte cuntun modul di pôc plui di 1200 regulis.

La disambiguazion e je une des fasis plui delicadis dai tradutôrs automatics, e chê dulà che al è plui facil cjatâ problemis che no àn une soluzion sigure e dulà che si produsin erôrs. Par chest, in plui dal slargjament cuantitatîf de base di dâts, i moduli di disambiguazion a son chei dulà che il lavôr di inzornament e perfezionament al devente plui impuartant, in consequence de osservazion dai erôrs fats di Jude.

- Trasferiment lessicâl e strutturâl: dopo fate la disambiguazion i lemis talians dai trinomis a vegnin voltâts in lemis furlans cussì, par esempi de secuencia `il_ART_MS porto_SM_MS` si passe ae secuencia `il_ART_MS puart_SM_MS`.

Cun di plui a son câs li che il trasferiment al da une soluzion componude di plui trinomis furlans rispjet a un sôl talian, par vie che a une peraule singule il furlan al fâs corrispuindi une locuzion (par esempi a *scendere* al pues corrispuindi *lâ jù*) o al contrari a puedin jessi soluzions di un sôl trinomi furlan rispjet a secuencis di plui trinomis talians (par esempi *buon senso* al varà di jessi voltât *sintiment*).

In plui di 2.000 câs, no je dome une sostituzion dal leme talian cun chel furlan tal trinomi, ma a cambiin ancje la categorie gramaticâl e/o morfologjiche, par esempi trasferint `mela_SF_FS` no si varà di rivâ a `miluç_SF_FS`, ma si varà di aplicâ il cambi di gjenar, tant

che se *mela* par talian al è feminin, *miluç* par furlan al è masculin e duncje il trasferiment strutturâl al proviodarà a dâ il trinomi coret miluç_SM_MS.

In chesj câs al coventarà ancje un trasferiment strutturâl: di fat se, par esempi, un sostantif al devente di feminin par talian a masculin par furlan, ancje articui, preposizions articoladis, adietifs o participis che i si riferissin a varan di lâi daûr cul stes cambiament par mantignî la concuordance. Une secunce tant che *la mela rossa* e passerà di *la_ART_FS mela_SF_FS rosso_ADI_FS* a *il_ART_MS miluç_SM_MS ros_ADI_MS*.

Tal trasferiment strutturâl a saran praticâts ancje altris intervents par fâ in maniere che la costruzion de frase si adati ae sintassi furlane dulà che si diferenzie di chê taliane: par esempi zontant i pronons clitics, cambiant di verp ausiliâr, cambiant la posizion des parts dal discors e v.i.

- Gjenerazion des formis furlanis: a chest pont, dai trinomis formâts di leme furlan - categorie gramaticâl - determinazion flessionâl si gjenerin lis formis furlanis, cussì par esempi di mangjâ_VTR_IP3 si gjenera la forme *mangje*. Pe gjenerazion des formis furlanis la Serling, cemût che si à za dite, e à dât un gnûf ordenament e e à slargjât la base di dâts dal Coretôr Ortografic Furlan, fasint in maniere che si rivi ae gjenerazion di dutis lis formis di passe 57.000 lemis furlans.
- Postgjenerazion: une volte gjenerât un test fat di une serie di formis furlanis la traduzion automatiche e je scuasi finide, te postgjenerazion a vegnin fats adaments su lis formis, par esempi l'articul *il* devant di peraulis che a tachin par vocâl al ven sostituît cu la forme cun elision *l'*: *il arbul* > *l'arbul*.

Ca sot si pues lâ daûr di une sintesi di dutis lis fasis cuntune frase esemplificative: *l'albicocca è buona come la pesca*

- tokenizazion: *l' | albicocca | è | buona | come | la | pesca*
- analisi morfologjiche: *l'[ella_PRONP_OA3f la_ART_FS il_ART_MS egli_PRONP_OA3m] albicocca[albicocca_SF_FS] è[essere_VINTR_IP3] buona[bua_SF_FSSUP1 buono_ADI_FS buono_SMF_FS] come[come_AV_UN come_CONIUN_UN] la[ella_PRONP_OA3f la_ART_FS la_SM_MS la_SM_MP] pesca[pesca_SF_FS pescare_VTR_IP3 pescare_VTR_IMP2]*
- disambiguazion gramaticâl: *la_ART_FS albicocca_SF_FS essere_VINTR_IP3 buono_ADI_FS come[come_AV_UN come_CONIUN_UN] la_ART_FS pesca_SF_FS*
- disambiguazion dal leme: *la_ART_FS albicocca_SF_FS essere_VINTR_IP3 buono_ADI_FS come_AV_UN la_ART_FS pesca_SF_FS*
- disambiguazion semantiche: *la_ART_FS albicocca_SF_FS essere_VINTR_IP3 buono_ADI_FS come_AV_UN la_ART_FS pesca2_SF_FS*
- trasferiment lessicâl: *la_ART_FS armelin_SM_FS jessi_VINTR_IP3 bon_ADI_FS tant_AV_UN che_CONIUN_UN la_ART_FS pierçul_SM_FS*
- trasferiment strutturâl: *il_ART_MS armelin_SM_MS lui_PRONP_SA3m jessi_VINTR_IP3 bon_ADI_MS tant_AV_UN che_CONIUN_UN il_ART_MS pierçul_SM_MS*
- gjenerazion formis furlanis: *il armelin al è bon tant che il pierçul*
- postgjenerazion: *l'armelin al è bon tant che il pierçul*

2.3.3 Limits di Jude 3

I tradutôrs automatics a àn bisugne di jessi provâts cun continuitât e i erôrs che a fasin a àn di jessi analizâts par inzornâ e perfezionâ in maniere progressive il program. Jude 3 al à za un funzionament suficient par jessi presentât e proponût al public, ma al à un margin di miorament ancjemò avonde larc.

In cheste fase i siei limits principâi a son:

- Numar di formis voltadis: cumò la base di dâts e je di plui di 58.000 lemis talians che si fletin in plui di 3.400.000 formis. Il numar di lemis al è za bon, ma al podarès rivâ a plui o mancul 80.000 lemis par garantî une cuviertidure otimâl ancje di tancj cjamps specialistics. La cressite previodude e je fra i 5.000 e i 10.000 lemis ad an.
- Tratament dai nons propriis: inte fase di cumò Jude al tradûs ducj i nons propriis cognossûts, sedi cuant che chest al è coret (*Alessandro Magno* > *Alessandri il Grant*) sedi cuant che chest nol è coret (*Alessandro Manzoni* > *Alessandri Manzons*). Il tratament dai nons propriis al è un pont delicât intune vore di programs di traduzion automatiche di cualsisedi lenghe: par ce che al tocje *Jude* si previôt di podê frontâ in maniere plui corete chest aspjet tra plui o mancul un an.
- Adataments complès: par otignî traduzions plui coretis e mancul “duris” a coventin adataments complès che a van di là de corrispondence di peraulis singulis. Par esempi *a portata di mano* > *sot man* al è un di chescj câs. Lis regulis che cumò a puartin a soluzions dal gjenar a frontin plui o mancul 5.000 câs (cualchi volte coletîfs), ma si previôt une lôr cressite continue.
- Disambiguazion: lis regulis di disambiguazion a puedin jessi perfezionadis ancjemò, in ogni câs al sarà tant difcil che ancje cuant che a sedin rivadis al nivel massim di definizion a puedin risolti in maniere corete plui dal 90% dai câs ambigus. Par altri a son ancje câs di ambiguitât cence soluzion, stant che plui di une soluzion proponude e risulde corete e la sielte e pues meti in dificoltât ancje une persone, se no je avonde informade sul contest. Par esempi la frase *so che vuole fare il pompiere* e podarès jessi voltade par furlan *o sai che al vûl fâ il pompîr*, ma ancje *o sai ce che al vûl fâ il pompîr*.
- Concuardancis sintatichis: il fat che Jude al funzioni daûr di un sisteme di “trasferiment superficial” al presupon che nol fasi une vere analisi sintatiche e semantiche dal test talian. Chest al compuarte che intal test furlan, soledut par vie dai cambis di gjenar o numar, o ben intal câs dai pronons clitics subiet, che par talian no son, la soluzion no sedi simpri juste. Chest al sucêt soledut cuant che il subiet al è lontan dal verp o cuant che la struture sintatiche e je involuçade o marcjade di elements semantics avonde clârs pe persone, ma no cognossûts dal program informatic. Par esempi di chestis trê frasis ca sot Jude al è in stât di fâ une traduzion corete des primis dôs, ma no de tierce:

la gatta miagola > *la gjate e sгнаule*

la gatta del vicino miagola > *la gjate dal vicin e sгнаule*

la gatta guarda il piccione e miagola > *la gjate e cjale il colomp e al sagnaule*

- Mantigniment dal formât dai tescj: no je stade ancjemò elaborade la version di Jude che e mantegni dal dut il formât dai documents (gjenar di caratar, inclinazion, dimension, justificazion, presince di imagjins, grafics, notis, coments e v.i.). Cumò il test al ven tradot cence conservâ nissun di chescj caratars.

2.3.4 Jude 3 in rêt

Dopo completade la fase di realizazion dai diviers modui di Jude 3, il program al è stât fat funzionâ sul web e al è stât metût a disposizion dal public. La version web e permet une prove a gratis fin a 500 peraulis, e e previôt la pussibilitât di registrâ un contrat a paiament.

Par fâ la traduzion automatiche si à di copiâ o scrivi intune barconete il test talian, limitât a 12.000 caratars par otimizâ la velocitât di esecuzion (mancul di 10 seconts), e inviâ il procès. Il test voltât al comparirà intune altre barconete a pene sot e si varà di fâ la sô revision. Par facilitâ cheste operazion, cuant che si clicarà suntune peraule di revisionâ e comparirà juste parsore une striche cu la frase originâl, in mût di podê fâ un confront imediât.

Une volte finide la traduzion automatiche, l'utent al pues simpri recuperâle, viodût che il program al ten in memorie il test originâl, il test risultât de traduzion automatiche e ancjechel risultât de revision dal utent.

Conclusion

La realizazion dal tradutôr automatic Jude e je passade par plui fasis, e à compuartât un invistiment privât di centenârs di oris di lavôr dilunc di plui di sîs agns, ma e je rivade al risultât di un program a trasferiment superficial che al permet a un tradutôr uman di lavorâ cuntun sparagn di timp di plui dal 50% rispjet ae traduzion manuâl, indotant cussì la lenghe furlane di un dai struments plui impuartants tal cjamp de linguistiche computazionâl.

Bibliografie

AA.VV. (2005) *An open-source shallow-transfer machine translation engine for the romance languages of Spain*

<http://www.dlsi.ua.es/~mlf/docum/corbibellot05p.pdf>

D.J.Arnold e al. (1994) *Machine Translation: An Introductory Guide*

<http://www.essex.ac.uk/linguistics/external/clmt/MTbook/HTML/book.html>

Eric Brill *A Simple Rule-Based Part of Speech Tagger*

<http://ucrel.lancs.ac.uk/acl/H/H92/H92-1022.pdf>

Eric Brill *Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging*

<http://www ldc.upenn.edu/acl/W/W95/W95-0101.pdf>

W.John Hutchins e Harold L. Somers (1992) *An introduction to machine translation*

<http://www.hutchinsweb.me.uk/IntroMT-TOC.htm>

Nicolas Stroppa e al. (2006) *Example-Based Machine Translation of the Basque Language*

<http://mt-archive.info/AMTA-2006-Stroppa.pdf>

Francis M. Tyers e Jacob Nordfalk *Shallow-transfer rule-based machine translation for Swedish to Danish*

<http://www.mt-archive.info/FreeRBMT-2009-Tyers.pdf>

www.serling.org