OpenLogos MT and the SAL Representation Language

Bernard Scott
Logos Institute
Tarpon Springs, USA
logos.institute@gmail.com

Anabela Barreiro
Logos Institute
Lisboa, Portugal
barreiro_anabela@hotmail.com

Abstract

This paper describes OpenLogos, a rule-driven machine translation system, and the syntactic-semantic taxonomy SAL that underlies this system. We illustrate how SAL addresses typical problems relating to source language analysis and target language synthesis. The adaptation of OpenLogos resources to a specific application concerning paraphrasing in Portuguese is also described here. References are provided for access to OpenLogos and to SAL.

1 Introduction

This paper seeks to describe the OpenLogos machine translation (MT) system and to illustrate how its semantic-syntactic representation language (SAL) contributes to effective MT, particularly with respect to the issues of complexity and ambiguity. We also show SAL's contributions to quality output. Finally, we explain the adaptation of OpenLogos resources and SAL to a new application for the Portuguese language.

2 System Architecture

OpenLogos system has the architecture of a pipeline, as illustrated in Figure 1, allowing for a modularized, incremental approach to source language analysis and target language synthesis. The representation language SAL (semanto-syntactic abstraction language) is the key to the pipeline process. Both input stream and rules are expressed in SAL and all interactions throughout the pipeline between rulebases and the input stream are in terms of SAL pattern matches. This homogeneity between rulebase and input stream has proven efficient in surmounting the classic problem of rulebase size as it relates to system performance. Because rules are in the form of SAL patterns, they are stored as indexed pattern dictionaries, allowing the rule-matching function to resemble ordinary lexical look-up. This characteristic of the system explains why, in the pursuit of high quality translation, OpenLogos rulebases have grown to many thousands of rules with minimal impact on system performance.

OpenLogos covers English source, with German, French, Spanish, Italian and Portuguese targets; and German source, with English, French and Italian targets. All software modules in the pipeline are language neutral. Language-specific information (lexicon, morphology, semantic-syntactic rule bases) is in the form of alpha-numeric data stored in tables in a relational database.

OpenLogos is a multi-target system. This means that once data files for a source language have been developed, data files for any number of targets can be linked thereto. However, the lexicon and associated morphology tables are fully multilingual, meaning dictionary entries for all languages serve both source and target purposes.

3 System Flow

The OpenLogos system flow that takes place in this pipeline is as follows: (1) A front-end software module strips away all document format codes, to be later re-applied to target output. (2) The LEX module presents the raw natural language (NL) input stream to the lexicon and converts the NL string to a SAL string. (3) Subsequent pipeline software modules seek to match the SAL input stream to corresponding SAL patterns in the rulebases. In this, SAL elements of the input stream serve as search arguments, in a similar way as NL words do in conventional dictionary look-up. For a match to occur on competing rules, rule constraints must be satisfied. Best match principles also apply, based on (a) degree of semantic specificity and (b) pattern length. (4) Upon match, a software module interprets the action component of the rule, driving source analysis in a bottom-up, deterministic fashion.

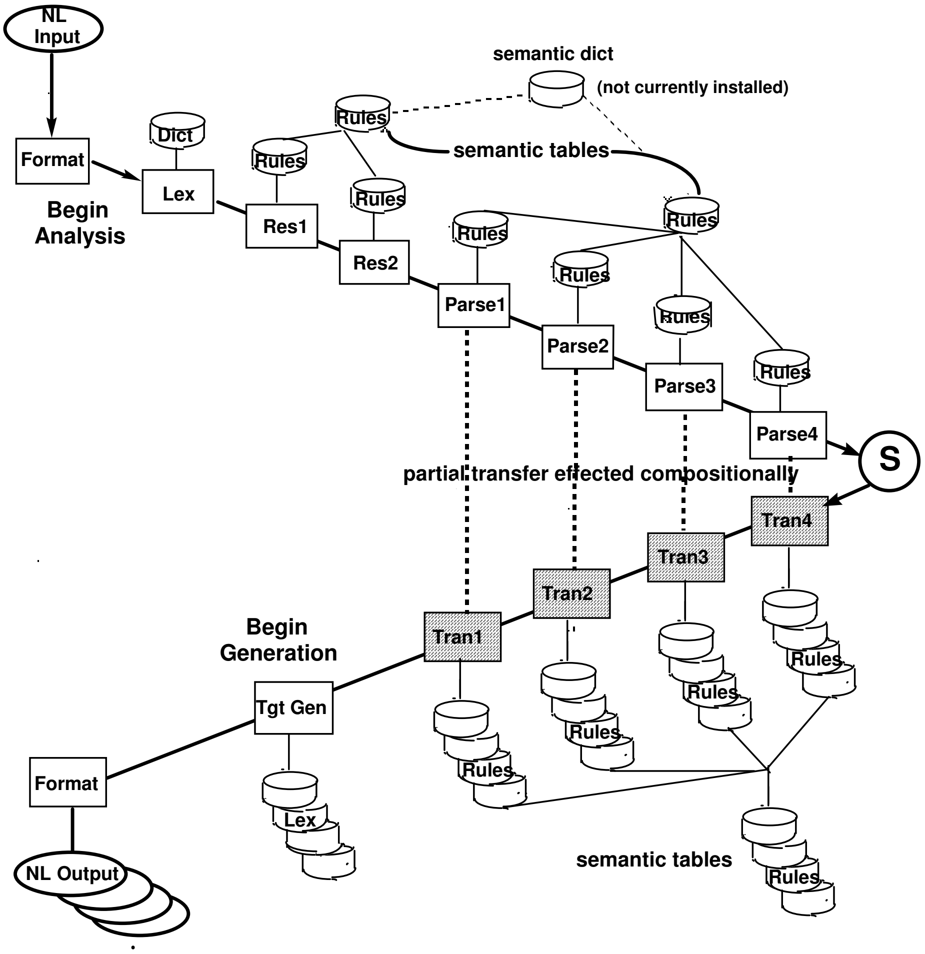


Figure 1: Modularized, Incremental Pipeline Architecture of OpenLogos

(5) Notations pertinent to target equivalences, both syntactic and semantic, are recorded when the analysis of each source constituent is completed, in contrastive linguistics (tree-to-tree) fashion. Target translation is generated upon completion of source analysis and source document format codes are re-applied to the target. In the pipeline process, the semantic tables, which both source and target rules access as appropriate, play a particularly interesting role, typically in resolving semantic ambiguities.

4 SAL Representation Language

The SAL taxonomy comprises approximately 1,000 elements (words), representing all parts of speech. As a higher order language to which NL maps, SAL is far richer than the symbols of syntax, but far leaner than natural language itself. Processing NL at this intermediate level is what has allowed OpenLogos to effect syntactic and semantic disambiguation without running into the complexity issues that have traditionally plagued MT (Scott 1998; 2001; 2003).

The SAL taxonomy has Supersets, Sets, and Subsets. For example, the SAL Superset for the noun *truck* is CO (Concrete), the SAL Set is COagen (Concrete/agent), and the SAL Subset is COvehic (Concrete agent/vehicle). SAL was designed to function like natural language at these more abstract levels, such that any NL string could be readily expressed by an equivalent SAL string, enabling developers and users alike to map easily from NL to SAL in lexical work. The relative simplicity in OpenLogos lexical work has helped it avoid the kinds of lexicography problems that have defeated some systems, such as TAUM, among others.

Semantics and syntax are treated in SAL as a continuum from literal string to word class (WC). For example, the input word *highchair* could be dealt with during pipeline analysis at any of the following representational levels:

Literal level:	*Highchair*
Head morpheme:	*Chair*
SAL Subset:	COsurf (Concrete/bearing surface)
SAL Set:	COfunc (Concrete/functional device)
SAL Superset:	CO (Concrete)
Word Class:	N

All SAL elements in both input stream and rulebase are characterized by the triplet: WC(Type; Form). In the input stream, Type always includes the element's Superset, Set and, if available, Subset (not all Sets have Subsets). Rules, on the other hand, specify Type at only one of the three levels, depending on the intended reach of the rule. For simplicity sake, we must ignore the 'Form' part of this triplet in the following discussion.

Although virtually all SAL codes at the Superset level are interlingual, SAL is not fully interlingual. It *is* interlingual, however, in the case of nouns and adverbs, and only somewhat less so in closed (grammatical) word classes, such as prepositions or determiners. Verbs and adjectives tend not to lend themselves to interlinguality.

Figure 2 shows in detail the Information Superset, of a total of 11 noun supersets, shown in Figure 3. The reader can obtain the entire SAL taxonomy at the Logos Archives website (referenced in Figure 3 caption).

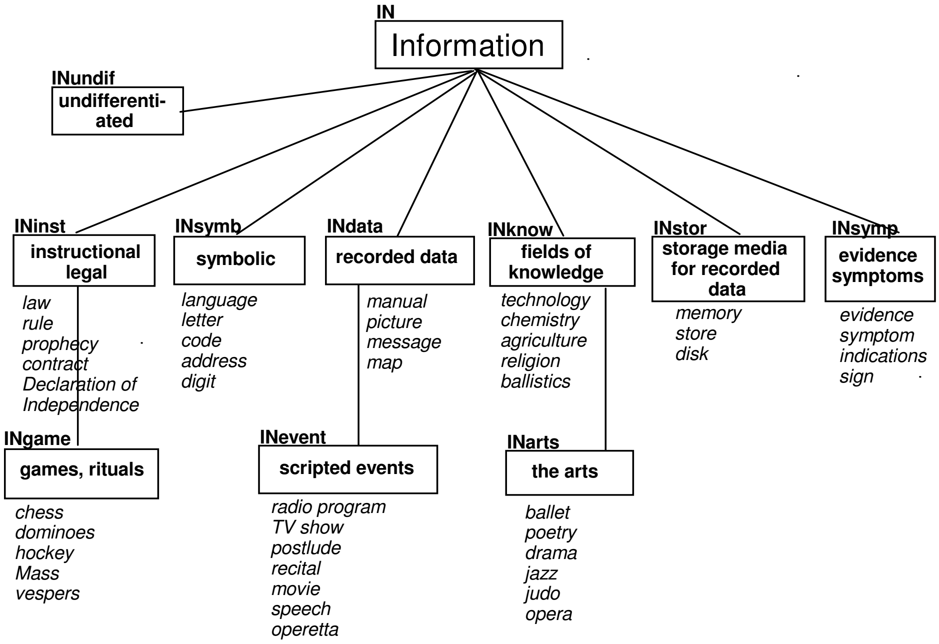


Figure 2: SAL Taxonomy for the Information Noun Superset with attending Sets and Subsets. Note that some Sets do not have Subsets.

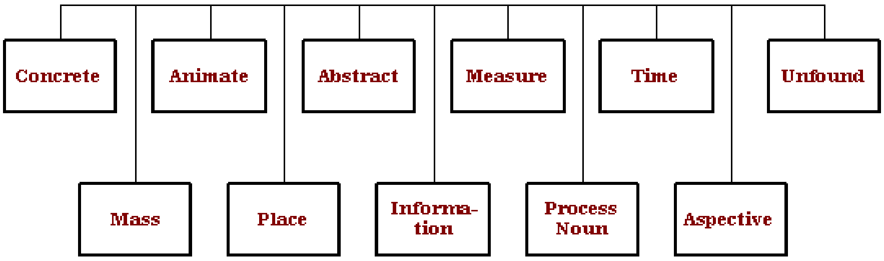


Figure 3: SAL Noun Supersets. Sets and Subsets within these Supersets are viewable at (http://logossystemarchives.homestead.com)

5 How SAL is Used in OpenLogos

Unlike with statistical MT, rule-based systems must analyze words morphologically, syntactically, and in advanced systems, semantically. Ambiguity of NL in each of these areas is the main difficulty with such analysis. In the following section, we show how SAL contributes to both syntactic and semantic disambiguation.

5.1 SAL Use in POS Disambiguation

In English, many verb particles are homographic with prepositions, and rule-based analysis has the task of resolving which part of speech applies in a given string. For example:

1) *He turned in the driveway.*
2) *He turned in the badge.*

In 1), *turned* is an intransitive verb and *in* is a preposition. SAL encodes *driveway* as PLpath (Place/path), enabling a SAL rule pattern comprising an intransitive motion verb, a preposition, and PLpath, thus to recognize *turn* as intransitive, and the ambiguous *in* as a preposition.

In 2), *turned* is a transitive verb and *in* is its particle satellite. SAL encodes *badge* as INdata (Information/recorded data), enabling a slightly more complex SAL rule to resolve *turn in* as a transitive verb with its verb particle. To effect this resolution, the rule must first send the pattern to the Semantic Table (see Figure 1) to determine that the transitive verb here does indeed take the particle. The added effect of this is that the particle/preposition ambiguity of the word *in* also gets resolved.

In Sentences 3) and 4), below, we see another type of disambiguation problem. In 3) *put* is an adjectival past particle. In 4), *put* is an active voice verb in the present tense. Rules that resolve these ambiguities depend upon SAL codes of both the subject and object. The effect of the POS ambiguity resolution can be observed in the German translations represented in 3') and 4').

3) *The emphasis put on the question was wrong.*
3') *Der Nachdruck, der auf die Frage gestellt wurde, wurde falsch.*
4) *John put on his hat.*
4') *John setzte seinen Hut auf.*

5.2 Structural Disambiguation: Resolving the Scope of Prepositions

The literal strings represented in 5) and 6) have similar syntactic structures but parse very differently, depending on the scope of their prepositions.

5) *to* citizens *of* Rome and friends
6) *to* citizens *of* Rome and environs

Seen purely as a syntactic string: [PREP1 N1 PREP2 N2 CONJ N3], 5) and 6) can be parsed either as 5') or 6'), depending on how the scope of PREP1 and PREP2 is interpreted.

5') PREP1 (N1 (PREP2 N2)) CONJ N3 →
PREP NP CONJ N
6') PREP1 (N1 (PREP2 (N2 CONJ N3))) →
PREP NP

The SAL elements that the LEX pipeline module substitutes for the NL nouns are as follows:

citizens, friends	→	ANdes (Animate/ human designation)
Rome	→	PLcity (Place/city)
environs	→	PLundif (Place/undifferentiated)

When NL strings 5) and 6) come out of the LEX pipeline module, each word in the string will have thus been replaced with its corresponding SAL element, shown in 5") and 6''). (Only SAL codes for the nouns in the input stream are shown.)

5") *to citizens of Rome and friends* →
to N(ANdes) *of* N(PL) *and* (ANdes) →
PREP1 (N1 (PREP2 N2)) CONJ N3
6") to citizens of Rome and environs →
to N(ANdes) *of* N(PL) *and* N(PL) →
PREP1 (N1 (PREP2 (N2 CONJ N3)))

SAL pattern rules in PARSE3 relating to preposition governance recognize semantic symmetries in these strings and thus achieve a proper parse. In 5"), because of the semantic symmetry, analysis interprets the preposition *to* as applying to both N1(*citizens*) and N3(*friends*). In 6"), the semantic symmetries (*Rome* and *environs)* are governed by the preposition *of*, producing a different parse.

The parses shown in 5") and 6") are accomplished incrementally by simple rule patterns firing over three of the pipeline modules (PARSE1-PARSE3). But in all cases, the rules involved are themselves

SAL patterns containing SAL noun codes that match, or do not match, those of the SAL input stream.

5.2.1 Resolving Subject/Object Ambiguity in German and English Source

In 7), the German subject/object word order must be reversed in the English translation, as evidenced in 7'):

7) *Dieses Garten liebt meine Mutter.*
7') *My mother loves this garden.*

The OpenLogos rules that effect this resolution and translation recognize *Mutter* rather than *Garten* as agent of the predicate *lieben*. The SAL Set code for *Mutter* is ANdes (Animate/human designation). The code for *Garten* is PLencl (Place/enclosed space).

In 8) and 9), the SAL Animate Subset code for *insects* is ANbugs, which is agentive by definition in OpenLogos. The SAL Mass Set code for *corn,* MAedib, is non-agentive. This distinction allows rules in the pipeline to render these output strings correct syntactically, notwithstanding some semantic and tense problems with verb treatment in the German translation. The French translation presents no such linguistic problem with verb handling.

8) *corn eating insects*
8') *les insectes qui mangent le maïs*
8") *Maisessen-Insekten*

9) *insects eating corn*
9') *les insectes qui mangent du maïs*
9") *Insekten, die Mais fraßen*

5.3.1 SAL Use in Resolving Morphological Agreement

In 10), the verb *effect* belongs to a SAL verb Set that must have as object a process noun. In 10), the noun *changes* is such a process noun and hence satisfies the requirement as the participle's object, causing *effected by digitalis* to be attached to *changes* morphologically. In 11), the SAL Set code for the verb *affect* offers no reason to attach *affected by digitalis* to anything but its left-adjacent noun, *tissue*, which is the default action. Note the effect of this on the morphology of the participles in the French translations.

10) *changes in tissue effected by digitalis.*

11) *changes in tissue affected by digitalis.*

10') *changements de tissu effectués par la digitaline.*
11') *changements de tissu affecté par la digitaline.*

5.4 SAL Use in Resolving Semantic Ambiguities

The verb *raise* in strings 12) to 14) is seen to be highly ambiguous. To determine its appropriate contextual meaning, a rule in PARSE3 sends the verb and its object (i.e., the head noun of the verb's NP object) to the Semantic Table for sense resolution. SAL codes for the verb's object provide the necessary clue to meaning and transfer. Note that in 12), the object *child* is coded ANdes (Animate/human designations). In 13) *corn* is coded MAedib (Mass/edible). In 14) *rent* is coded MEabs (Measurement/abstract concepts measured by units, e.g., dollars, euros, etc.). This results in three different translations for the verb in French.

12) *raise a child* →
V('*raise*') N(ANdes) → *élever. . .*
13) *raise corn* → V('*raise*') N(MAedib) → *cultiver. . .*
14) *raise the rent* → V('*raise*') N(MEabs) → *augmenter. . .*

The single, simple "deep structure" rule from the Semantic Table in 14), is shown below, in 15) to 18), as having the ability to deal not only with the sense of *raise* as a verb, but also with all its derivative "surface structure" forms, both in the English source and in the French target.

15) *he raised the rent* → *il a augmenté le loyer*
16) *the raising of the rent* →
l'augmetation du loyer
17) *the rent, raised by . . .* → *le loyer, augmenté de . . .*
18) *a rent raise* → *une augmentation de loyer*

These semantic transfers are effected by a single rule in PARSE3, at a relatively late stage of parse tree construction. The rule has the simple, purely syntactic pattern as follows: V NP. The action portion of this rule sends these two elements to the Semantic

Table for matches on semantic rules as shown above in 12)-14).

The simplicity and power of this pipeline architecture, coupled with SAL, represent the principal, distinguishing features of OpenLogos and accounts for the generally good quality of OpenLogos output.

6 Quality of OpenLogos Output

OpenLogos is a open-source version of the high-end Logos System (1970-2000), a commercial product used by scores of commercial users in Europe and North America. Users report that the system was capable of producing output requiring only a modest degree of post-editing (Cremers, 1993). Of course, not all output was consistently of uniformly high quality, but output approaching human quality was not uncommon where source documents were mindfully written and lexical work was well done.

In 19) and 20), we offer two short examples of translation quality, showing target voice transformations that we trust will illustrate the benefit of pipeline architecture interacting with a semantic-syntactic knowledge base expressed in SAL.

19) *The situation was alluded to by my friend in his letter.*

19')*Mon ami a fait allusion à la situation dans sa lettre.*

20) *The situation was alluded to in their letter.*

20') *On a fait allusion à la situation dans leur lettre.*

7 Open-Source Availability of SAL and OpenLogos Resources

OpenLogos is an open-source copy of the Logos System, implemented by DFKI and downloadable from the DFKI website at: http://logos-os.dfki.de. All lexical data for English, German, French, Italian, Spanish and Portuguese are available and accessible in a relational database. Also thousands of semantic rules in the Semantic Table (Semtab) are also available. DFKI offers this open-source release of the Logos System on the Linux platform using PostgreSQL as database.

SAL in its entirety may also be accessed at the Logos System Archives website at: http://logossystemarchives.homestead.com. This Logos archive site also provides expansive overviews of SAL and summaries of the internal processes of OpenLogos, as well as technical papers relating to the Logos Model. This legacy website does not provide assistance pertaining to the operation or downloading of OpenLogos. For that consult DFKI.

8 Using OpenLogos for New Applications

Using the NooJ development environment (Silberztein, 2006; 2008), the OpenLogos linguistic database at DFKI has been exploited to produce new applications: (1) ReEscreve (in English, ReWriter), a new paraphrasing system, currently available only for the Portuguese language, described in Barreiro (2008b; 2009), and (2) ParaMT, a prototype of a model for MT using paraphrases (Portuguese-English), described in Barreiro (2007; 2008b). In Section 8.1, we briefly describe the ReEscreve application. ParaMT is still at an initial state. It is a tool that uses similar methodology and mechanisms to those employed by ReEscreve. However, since ParaMT uses bilingual data, it is directly applicable to MT.

8.1 ReEscreve

ReEscreve is a language composition tool that helps authors to improve text quality and consistency by providing paraphrases that can substitute the content of existing text, standardize text style or optimize meaning. ReEscreve uses linguistically based automated paraphrasing and text-editing mechanisms to help users with their writing needs by providing suggestions for customized text authoring. It also generates word and phrasal usage data to help guide decision-making.

The linguistic resources used by ReEscreve are Port4NooJ (Barreiro, 2008a), currently at version 2.0. Port4NooJ contains electronic dictionaries (derived originally from OpenLogos English-Portuguese dictionary data, adapted and enhanced with new morpho-syntactic and semantic properties); new inflectional and derivational rules, and transformational grammars.

8.1.1 Paraphrasing Capabilities

ReEscreve currently transforms several categories of linguistic phenomena: (1) verbs into their synonyms, such as *ensinar* (to teach) into *leccionar* (to lecture); (2) different types of multiword adverbs into one word adverbs, such as *de um modo interactivo* (in an interactive way) into *interactivamente* (interactively), and vice-versa; (3) relative clauses into possessives, such as *o papel que a Europa tem/desempenha* (the role that Europe has/plays) into *o papel da Europa* (the role of Europe), and vice-versa; (4) relative clauses into adjectives, such as *os textos que foram escritos por* (the texts that were written by) into *os textos escritos por* (the texts written by) and vice-versa; (5) passives into actives, such as *a maçã foi comida pelo homem* (the apple was eaten by the man) into *o homem comeu a maçã* (the man ate the apple), and vice-versa; (6) different types of support verb constructions into verbs, such *as dar um abraço a* (to give a hug to) into *abraçar* (to hug); *causar incómodo* (lit. to cause disturbance) into *incomodar* (to disturb), or *entrar em contacto com* (lit. to enter into contact with) into *contactar* (to contact), and vice-versa; (7) support verb constructions into stylistic variants, such as *fazer uma viagem* (to go on a trip) into *efectuar uma viagem* (lit. to perform a trip), or *fazer impressão* (to make impression) into *causar impressão* (lit. to cause impression), and vice-versa; (8) stylistic variants of support verb constructions into verbs, such as *realizar um negócio* (to make business) into *negociar* (to negotiate), and vice-versa; (9) aspectual constructions into verbs, such as *começar um ataque* (to launch an attack) into *atacar* (to attack), and vice-versa; (10) predicate (resultative) adjective constructions into verbs, such as *estar/ficar cansado* (to be tired) into *cansar-se* (lit. to tire), and vice-versa. Refinement of the current dictionary's morpho-syntactic and semantic properties and grammars takes place and new grammars (and therefore paraphrasing capabilities) are being developed. These grammars complement Open-Logos Semtab rules, enabling the integration of paraphrases and the possibility to select more natural translations.

8.1.2 Web Interface/User Interaction

The ReEscreve web interface incorporates web technologies like Ajax to NooJ. Port4NooJ resources were developed in the NooJ environment, but, in principle anyone with NooJ could also, independently, use the grammars and lexicons in their own environment or even adapt them to a different environment. See Port4NooJ webpage for more details.

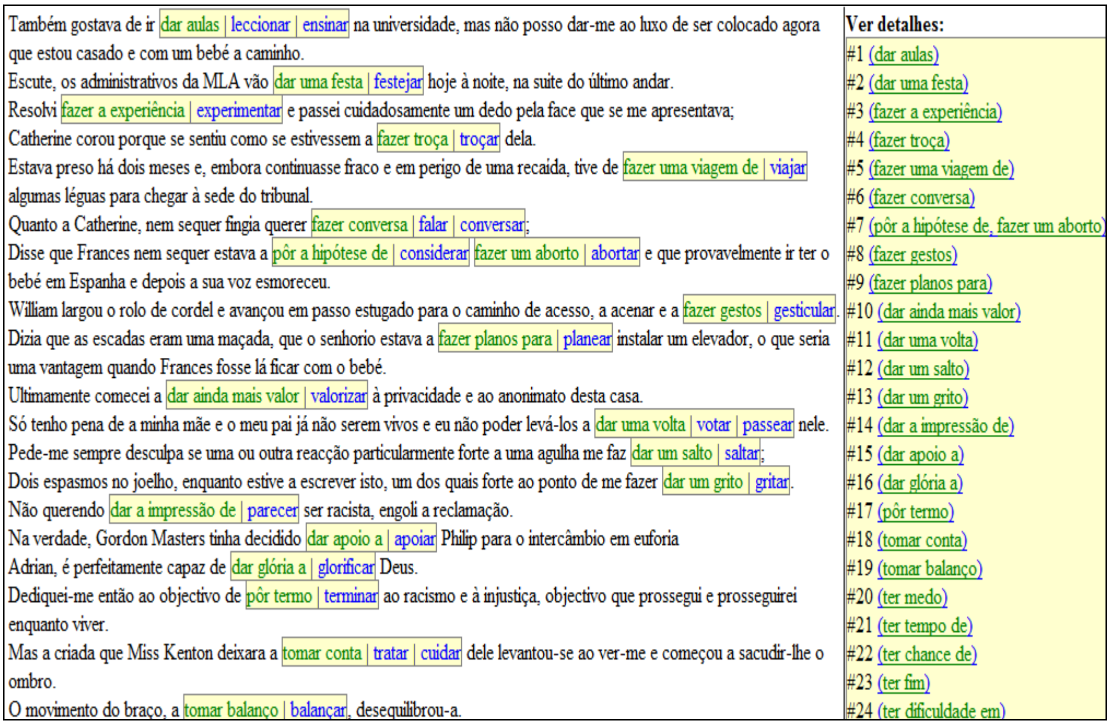
ReEscreve recognizes and converts certain words, expressions or phrases into semantically equivalent words or expressions (synonyms or paraphrasing capabilities), following three main steps: (1) recognition of a word, phrase or expression in a text and annotation; (2) annotated expressions are matched against a paraphrase database (developed in NooJ) and the corresponding candidates are retrieved; (3) suggestions are presented to the user side-by-side with the original expression.

The user operates ReEscreve interactively or in fully automatic mode. The user submits either text or a file containing text, and is shown where text needs editing because of wordiness, lack in clarity, or imprecision. Text changes take place at the word and at the phrase level (multiword expressions).

When used with terminological and domain specific dictionaries, ReEscreve helps with technical writing and prepares texts for machine translation. Pre-editing and linguistic quality control contributes to better texts and better translations. Used interactively, ReEscreve constitutes a learning tool for the native and especially for the second language learner user.

Figure 4 illustrates the interactive use of ReEscreve, where alternative suggestions are automatically placed in parallel to the user's original expressions. Upon deciding which expression better suits the objectives of the text, users can click on that expression and interactively rewrite their text online, one option at a time, or define a style and request automated changes to suit that style in the whole text.

ReEscreve is publicly available at: http://poloclup.linguateca.pt/ReEscreve/.



Também gostava de ir dar aulas | leccionar | ensinar na universidade, mas não posso dar-me ao luxo de ser colocado agora que estou casado e com um bebé a caminho.
Escute, os administrativos da MLA vão dar uma festa | festejar hoje à noite, na suite do último andar.
Resolvi fazer a experiência | experimentar e passei cuidadosamente um dedo pela face que se me apresentava;
Catherine corou porque se sentiu como se estivessem a fazer troça | troçar dela.
Estava preso há dois meses e, embora continuasse fraco e em perigo de uma recaída, tive de fazer uma viagem de | viajar algumas léguas para chegar à sede do tribunal.
Quanto a Catherine, nem sequer fingia querer fazer conversa | falar | conversar;
Disse que Frances nem sequer estava a pôr a hipótese de | considerar fazer um aborto | abortar e que provavelmente ir ter o bebé em Espanha e depois a sua voz esmoreceu.
William largou o rolo de cordel e avançou em passo estugado para o caminho de acesso, a acenar e a fazer gestos | gesticular.
Dizia que as escadas eram uma maçada, que o senhorio estava a fazer planos para | planear instalar um elevador, o que seria uma vantagem quando Frances fosse lá ficar com o bebé.
Ultimamente comecei a dar ainda mais valor | valorizar à privacidade e ao anonimato desta casa.
Só tenho pena de a minha mãe e o meu pai já não serem vivos e eu não poder levá-los a dar uma volta | votar | passear nele.
Pede-me sempre desculpa se uma ou outra reacção particularmente forte a uma agulha me faz dar um salto | saltar;
Dois espasmos no joelho, enquanto estive a escrever isto, um dos quais forte ao ponto de me fazer dar um grito | gritar.
Não querendo dar a impressão de | parecer ser racista, engoli a reclamação.
Na verdade, Gordon Masters tinha decidido dar apoio a | apoiar Philip para o intercâmbio em euforia
Adrian, é perfeitamente capaz de dar glória a | glorificar Deus.
Dediquei-me então ao objectivo de pôr termo | terminar ao racismo e à injustiça, objectivo que prossegui e prosseguirei enquanto viver.
Mas a criada que Miss Kenton deixara a tomar conta | tratar | cuidar dele levantou-se ao ver-me e começou a sacudir-lhe o ombro.
O movimento do braço, a tomar balanço | balançar, desequilibrou-a.

Ver detalhes:
#1 (dar aulas)
#2 (dar uma festa)
#3 (fazer a experiência)
#4 (fazer troça)
#5 (fazer uma viagem de)
#6 (fazer conversa)
#7 (pôr a hipótese de, fazer um aborto)
#8 (fazer gestos)
#9 (fazer planos para)
#10 (dar ainda mais valor)
#11 (dar uma volta)
#12 (dar um salto)
#13 (dar um grito)
#14 (dar a impressão de)
#15 (dar apoio a)
#16 (dar glória a)
#17 (pôr termo)
#18 (tomar conta)
#19 (tomar balanço)
#20 (ter medo)
#21 (ter tempo de)
#22 (ter chance de)
#23 (ter fim)
#24 (ter dificuldade em)

Figure 4: ReEscreve's suggestions for sample sentences with support verb constructions (Interactive use)

9 Conclusions

This paper has described the rule-driven OpenLogos machine translation system and its SAL representation language. The paper has sought to illustrate SAL's advantages for translation quality. SAL's availability to the open-source MT community was also described. The paper concludes with a description of an application that draws upon OpenLogos resources.

References

Barreiro, Anabela (2007). "ParaMT: a Paraphraser for Machine Translation". In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (eds.), *Computational Processing of the Portuguese Language, 8th International Conference, Proeed-ings (PROPOR 2008)* Vol. 5190, (Aveiro, Portugal, 8-10 de Setembro de 2008), Springer Verlag, pp. 202-211.

Barreiro, Anabela (2008a). "Port4NooJ Linguistic Resources Overview". http://poloclup.linguateca.pt/Port4NooJ/ResourcesOverview.pdf

Barreiro, Anabela (2008b). "Make it simple with paraphrases." PhD dissertation. Faculdade de Letras da Universidade do Porto. Porto, Portugal.

Anabela Barreiro & Luís Miguel Cabral (2009). "ReEscreve: a translator-friendly multi-purpose paraphrasing software tool". In Marie-Josée Goulet, Christiane Melançon, Alain Désilets & Elliott Macklovitch (eds.), *Proceedings of the Work-shop Beyond Translation Memories: New Tools for Translators*, The Twelfth Machine Translation Summit (Château Laurier, Ottawa, Ontario, Canada, 29 August 2009), pp. 1-8.

Cremers, Lou (1993). "Better, Cheaper, Faster Translation at Océ Technologies." Euromap – HTL Case Study. http://syntactic.com/page9/page9.html

Scott, Bernard (1998). "Linguistic and Computational Motivations for the Logos Machine Translation System." Technical Paper, Logos Corporation. http://www.iai.unisb.de/iaien/iaiwp/p11/index.html

Scott, Bernard (2001). Logos System: Principles and Motivations. http://logossystemarchives.homestead.com

Scott, Bernard (2003). "The Logos Model: An Historical Perspective." In: *Machine Translation* 18, 1-72.

Silberztein, Max (2006). NooJ manual. Downloadable from: http://www.nooj4nlp.net.

Silberztein, Max (2008). Complex Annotations with NooJ. In *Proceedings of the 2007 International NooJ Conference*. Cambridge Scholars Publishing: Newcastle, 2008.