

First International Workshop on Free/Open-Source Rule-Based Machine Translation

2nd and 3rd November 2009
Universitat d'Alacant, Alacant - Spain



Bernard Scott
logos.institute@gmail.com

Anabela Barreiro
barreiro_anabela@hotmail.com

Presentation Outline

- Background on OpenLogos
- System Pipeline Architecture
- SAL Representation Language
- Classic problems with rule-driven systems
- How SAL benefits translation
- Exploiting OpenLogos MT for new applications
- Free resources of OpenLogos

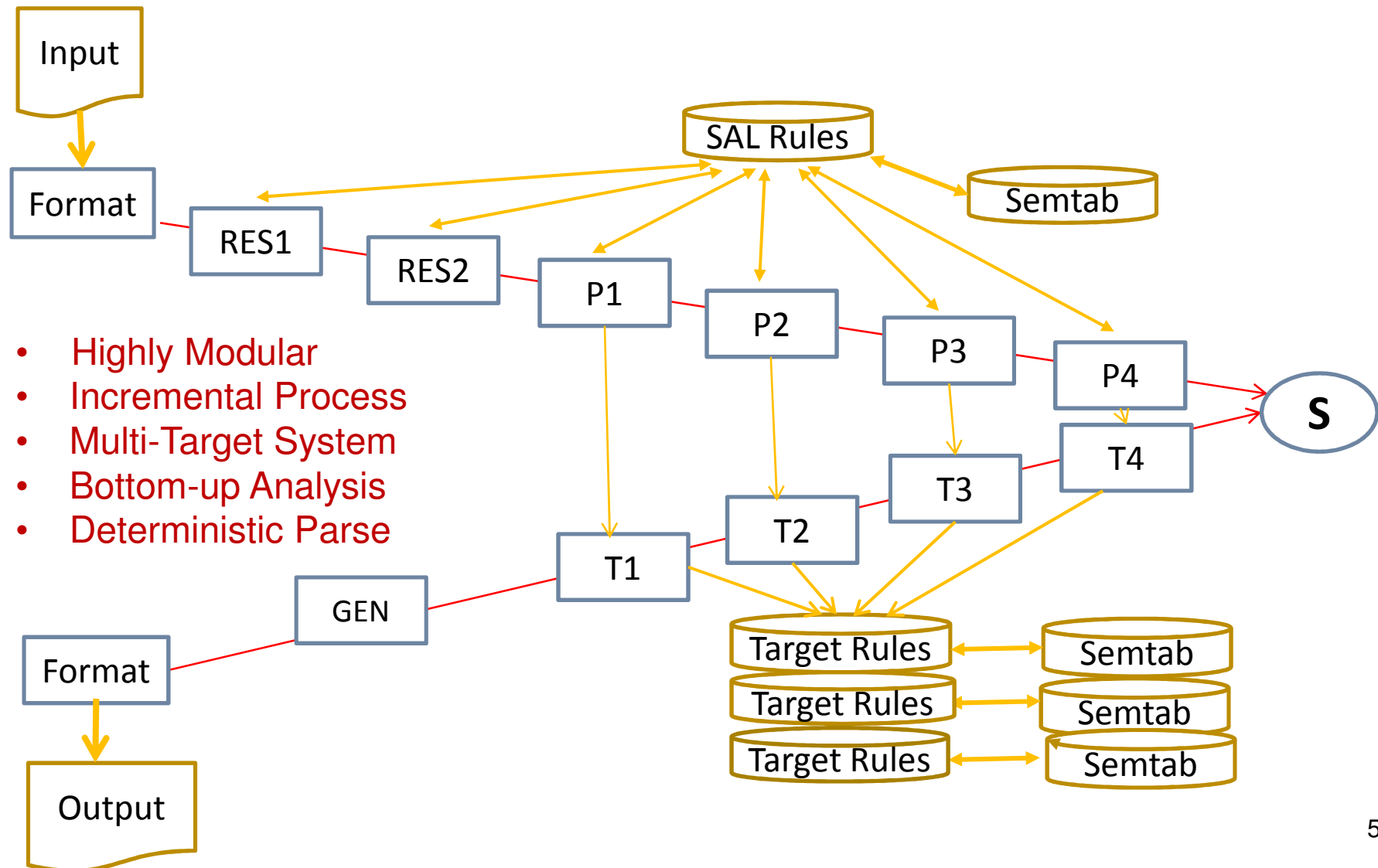
Background to OpenLogos

- Adapted from Logos System (1970-2000) by DFKI
 - Developed in US, Germany, Italy
 - 25-100 development staff for 30 years
- Language pairs: E-G, E-F, E-S, E-I, E-P
G-E, G-F, G-I
- Industrial strength MT used in 12 countries

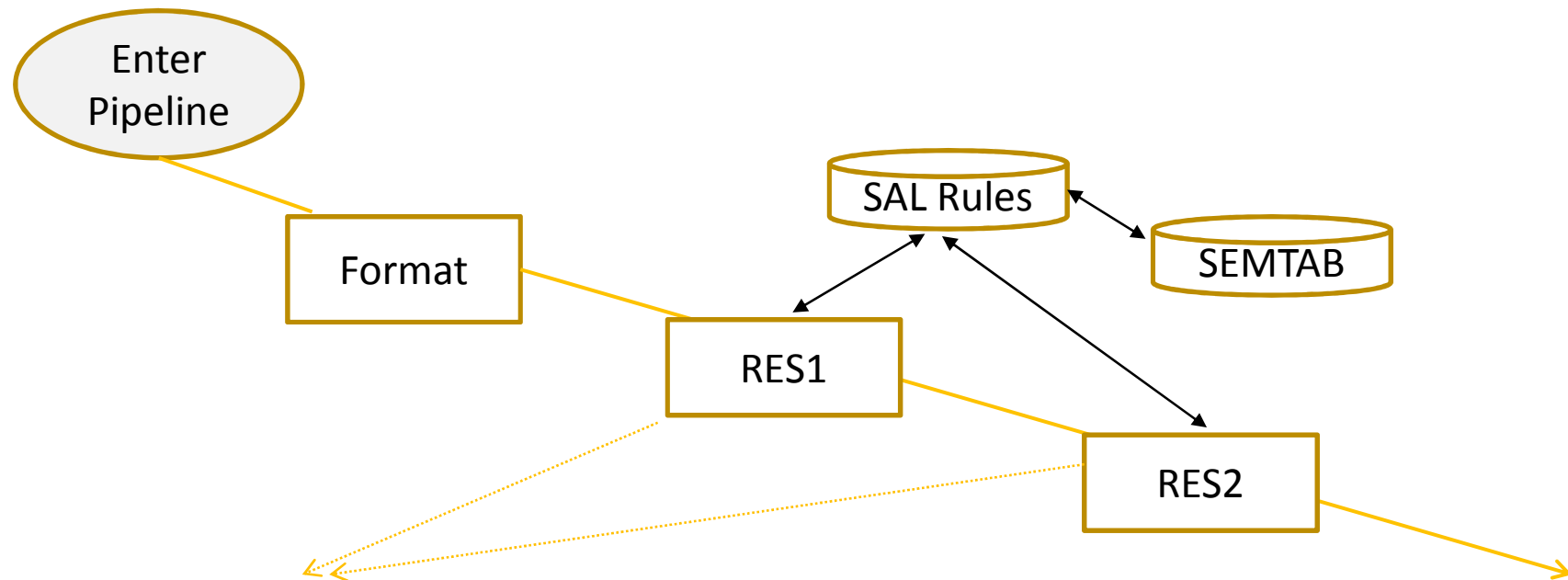
OpenLogos Characteristics

- Multi-target System
- Pipeline Architecture
- Language-neutral Software
 - All linguistic knowledge is in form of data, stored in a relational database
- SAL representation Language
 - Semanto-Syntactic Abstraction Language
- Semantic Processing
 - Thousands of transformation rules in Semantic Table (= Semtab)

OpenLogos Pipeline Architecture

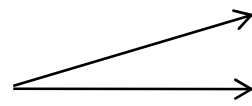


Incremental Source Analysis - 1



Clausal Segmentation

Homograph Resolution

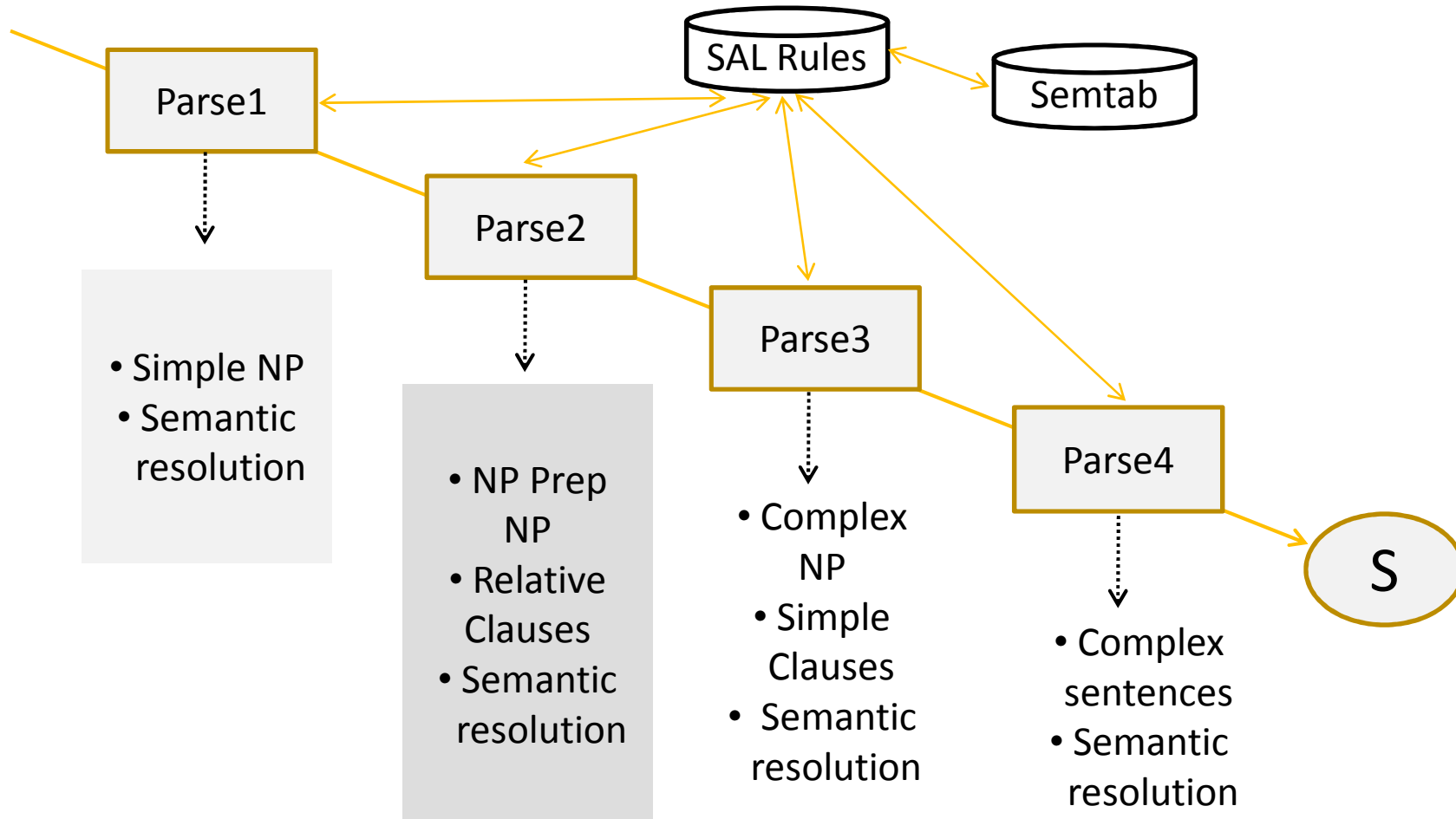


ways of cooking lentils - V

types of cooking utensils - ADJ

Deterministic parsing requires that all ambiguous PoS be resolved

Incremental Source Analysis - 2



E.g: a book **on** the presidency
on = **about**; **concerning**
≠ a book **on** the table

SAL Representation Language

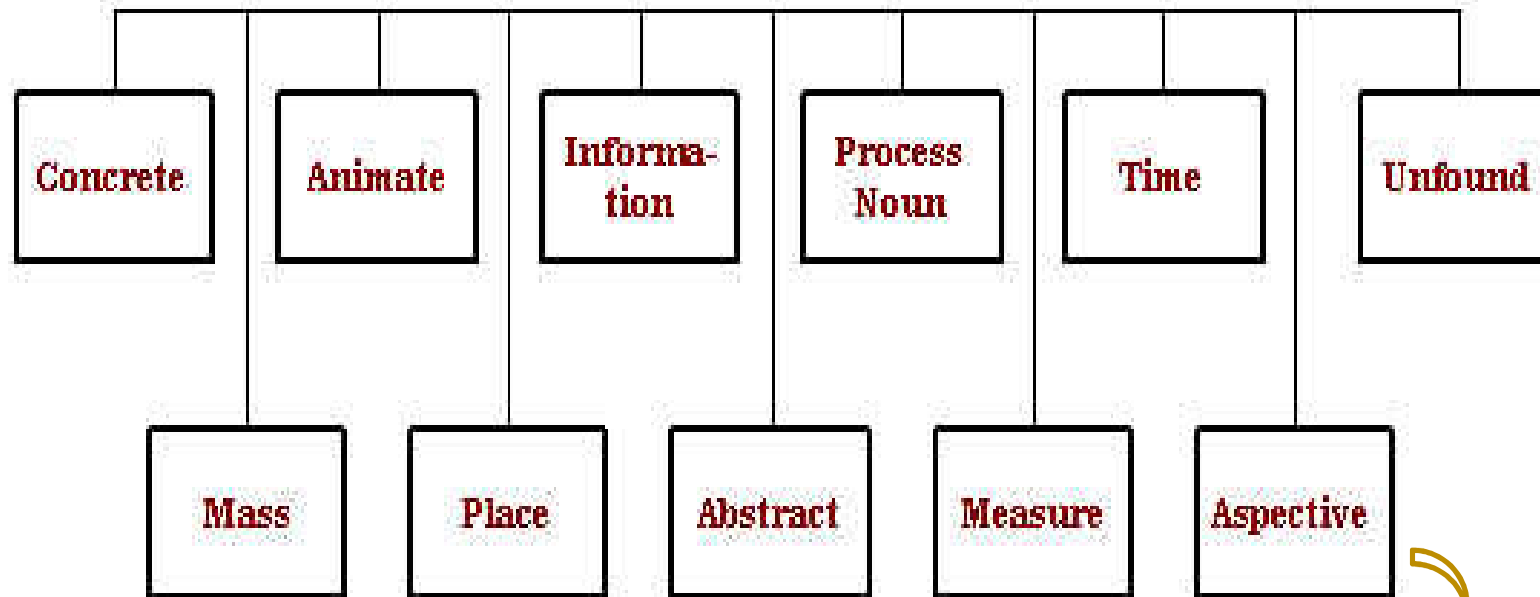
SAL stands for **S**emanto-syntactic **A**bstraction **L**anguage

- A Semanto-Syntactic Taxonomy: **Supersets** / **Sets** / **Subsets**
- Semantic-Syntactic continuum from NL word to Word Class
 - Literal word: *airport*
 - Head morph: *port*
 - **SAL Subset**: **agfunc** (agentive functional location)
 - **SAL Set**: **func** (functional location)
 - **SAL Superset**: **PL** (place)
 - Word Class: N

Both **Pipeline Input Stream** and **Rulebases** are in **SAL**

SAL Noun Supersets

SAL Noun Supersets



Developed:

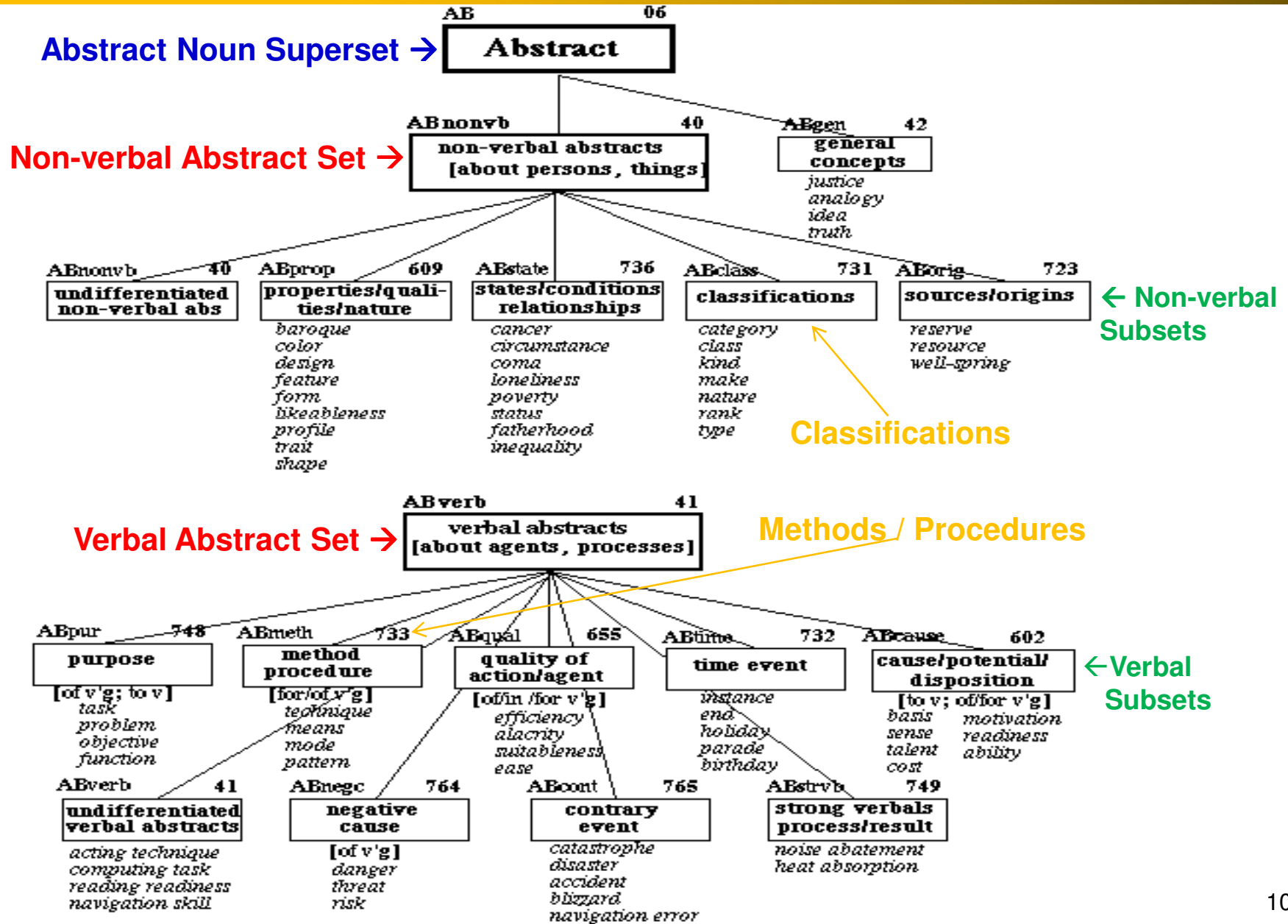
- inductively
- by trial and error
- over a period of years
- by the development team

E.g: **two pieces of cake**

NP parse must have:

- Plural morphology of pieces
- Semantics of cake

Abstract Noun Taxonomy



Use of SAL Codes to Resolve Homographs

Is the word *cooking* a verb or an adjective?

ways of cooking lentils

types of cooking utensils

<i>ways</i>	→ N(AB/ method)	→ parser verb bias
<i>types</i>	→ N(AB/ class)	→ non-verb bias

The SAL code **N(AB/method)** in the rule matches on a similar code in the SAL input stream. The effect of such a match is to resolve *cooking* as a **verb**

SAL contributes to the resolution of the homograph

What SAL Rules Look Like

Rules Have Five Components

- **SAL Pattern**

- PARSE2 Example: N(**IN/data**;u) Prep(“on”;u) N(u;u) (a book **on** the presidency)

- **Constraints**

- Match only if conditions are true or false

- **Source Actions**

- RES Rulebase: Resolves syntactic ambiguity
- PARSE Rulebase: Creates parse tree
- SEMTAB Rules: Effect semantic disambiguation

- **Target Action (optional)**

- Effect syntactic and/or semantic transfer

- **Comment Line**

- PARSE2 Example: NP(info) Prep(“**on**”) NP → N1 “**about**” N2
E.g., **book on** political **satire** → **book about**



Classic Problem of RBMT

- Complexity
 - Logic Saturation
 - Performance Degradation
 - Difficult Maintainability
 - System Improvability Stasis
- Ambiguity
 - Quality/Accuracy of Output – depends on effective disambiguation
- Classic Dilemma of the Developer
 - Reduce rulebase size to relieve complexity weakens disambiguation
 - Increase rulebase size to address ambiguities increases complexity

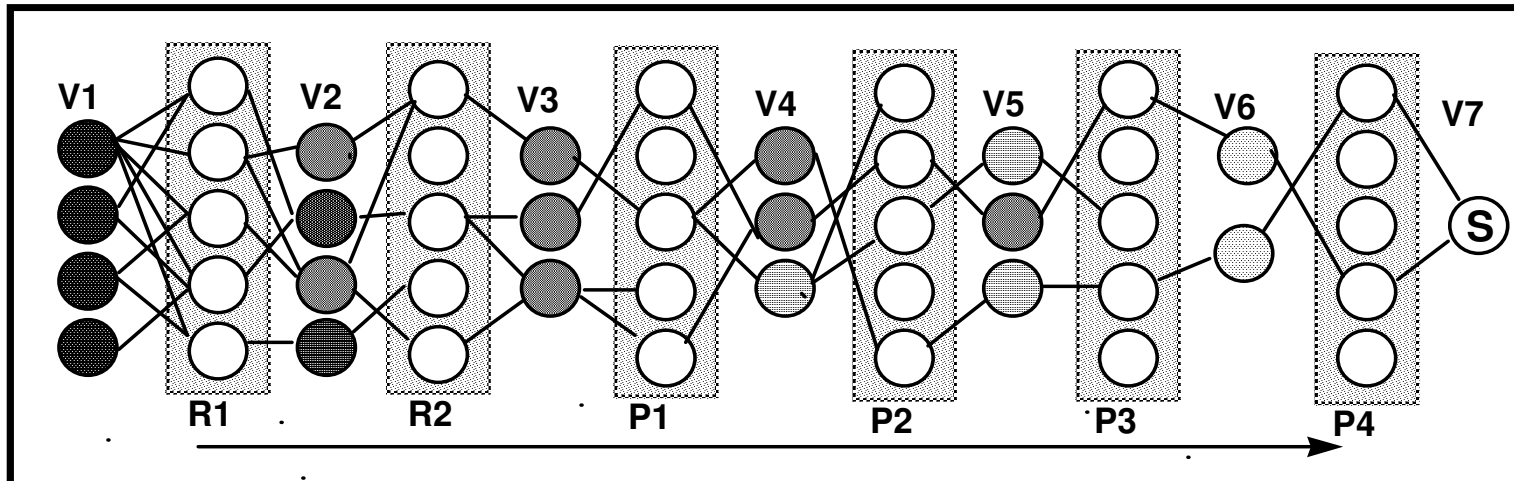
How OpenLogos Addresses Complexity and Ambiguity

- Complexity

- Rules and input stream are both expressed as **SAL patterns**
- Homogeneous **apples-to-apples matching**
- SAL input stream serves as **search argument** to SAL rulebase
- Rules are SAL patterns stored in an **indexed pattern dictionary**
- **No limit on rule size, no impact on performance**
- **Rules are self organizing**, like a dictionary
- **Rulebase is easy to maintain**

How Rules Are Applied

Metaphor: biological neural net



As the analysis progresses:

1- cells become fewer (abstract nature of the parse)

2- vectors become lighter (semantic disambiguation)

- Vectors labeled V1-V6 = SAL input stream of the pipeline
- Cells in input vectors = SAL elements/words to which the NL input stream has been converted
- In this network, R1 through P4 = hidden layers containing SAL rules
- R1 represents RES1, P1 represents Parse1 and so on.
- Each hidden layer contains between 2-4 thousand rules, organized by their SAL pattern, as in a dictionary.

How OpenLogos Addresses Complexity and Ambiguity

- Ambiguity
 - Syntactic Homograph Resolution
 - Scoping of adjectives, prepositions
 - Polysemy

Resolution of Polysemy in OpenLogos

Key to Polysemy Handling:

1. SAL Representation Language
2. Semantic Table

NL String

raise a child

raise corn

raise the rent

Sem. Table Rule

→ V('raise') N(**ANdes**)

→ V('raise') N(**MAedib**)

→ V('raise') N(**MEabs**)

French Transfer

→ *élever...*

→ *cultiver...*

→ *augmenter...*

Deep Structure Rules of Semantic Table

A Single Deep-Structure Rule
Matches Multiple Surface-Structures
And Produces Appropriate Target Transfers

<i>he <u>raised</u> the rent</i>	→	<i>il a <u>augmenté</u> le loyer</i>	V+Object
<i>the <u>raising</u> of the rent</i>	→	<i>l'<u>augmentation</u> du loyer</i>	Gerund
<i>the rent, <u>raised</u> by ...</i>	→	<i>le loyer, <u>augmenté</u> de ...</i>	Part. ADJ
<i>a rent <u>raise</u></i>	→	<i>une <u>augmentation</u> de loyer</i>	Noun

How SAL Benefits Translation

Examples showing
target voice transformations

EN passive voice >>> FR active voice

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

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

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

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

Contribution of OpenLogos Resources for New NLP Applications

● ReEscreve (ReWriter)

- Language composition tool
- Helps authors to improve text quality and consistency
- Provides paraphrases that can substitute the content of existing text, standardize text style or optimize meaning
- Based on NooJ's technology (<http://ww.nooj4nlp.net/>)
- Publicly available at: <http://www.linguateca.pt/ReEscreve/>

● ParaMT

- Bilingual/multilingual paraphraser (translator prototype)
- Uses similar methodology to that employed by ReEscreve
- Uses bilingual data
- Directly applicable to MT

● Corpógrafo

- Multilingual corpora management tool
- Available at: <http://www.linguateca.pt/corpografo/>

Applications Employing OpenLogos Resources

OpenLogos EN-PT dictionary data were adapted and enhanced with new properties (derivational, etc.) to create a new resource: Port4NooJ (<http://www.linguateca.pt/Repositorio/Port4NooJ/>). The new applications created for Portuguese NLP use Port4NooJ.

ReEscreve: Suggestions for Text Rewriting

Também gostava de ir [dar aulas](#) | [leccionar](#) | [ensinar](#) na universidade, mas não posso dar-me a isso 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, para ver se 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.

Paraphrases of SVC
presented by
ReEscreve's
paraphrasing system

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](#))

ReEscreve: a Rewritten Text

Também gostava de **leccionar** 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** hoje à noite, na suite do último andar.

Resolvi **testar** e passei cuidadosamente um dedo pela face que se me apresentava;

Catherine corou porque se sentiu como se estivessem a **fazer troça** dela.

Estava preso há dois meses e, embora continuasse fraco e em perigo de uma recaída, tive de **viajar** algumas léguas para chegar à sede do tribunal.

Quanto a Catherine, nem sequer fingia querer **fazer conversa**;

Disse que Frances nem sequer estava a **pensar interromper a gravidez** 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 **gesticular**.

Dizia que as escadas eram uma maçada, que o senhorio estava a **planear** instalar um elevador, o que seria uma vantagem quando Frances fosse lá ficar com o bebé.

Ultimamente comecei a **apreciar** a 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 **passear** nele.

Pede-me sempre desculpa se uma ou outra reacção particularmente forte a uma agulha me faz **saltar**;

Dois espasmos no joelho, enquanto estive a escrever isto, um dos quais forte ao ponto de me fazer **gritar**.

Não querendo **parecer** ser racista, engoli a reclamação.

Na verdade, Gordon Masters tinha decidido **apoiar** Philip para o intercâmbio em euforia

Adrian, é perfeitamente capaz de **dar glória a Deus**.

Dediquei-me então ao objectivo de **pôr termo** ao racismo e à injustiça, objectivo que prossegui e prosseguirei enquanto viver.

Mas a criada que Miss Kenton deixara a **tomar conta** dele levantou-se ao ver-me e começou a sacudir-lhe o ombro.

O movimento do braço, a **balançar**, desequilibrou-a.

não há de que **ter medo**.

Sai da loja antes de ele **ter tempo de** chamar a brigada de narcóticos.

Text rewritten based
on the user's
preferences

Users can suggest
new expressions!

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 [\(ter dificuldade em\)](#)

ParaMT: a Paraphraser Applicable to MT

PT support verb construction

a fazer um estágio para dar aulas de tutor
a fazer um estágio para dar aulas de lecture
a fazer um estágio para dar aulas de teach
ra que Edward pudesse dar inicio /start
ra que Edward pudesse dar inicio /initiate
ra que Edward pudesse dar inicio /commence
ra que Edward pudesse dar inicio /begin
os no Natal.» -- Estou a dar aulas para/tutor
os no Natal.» -- Estou a dar aulas para/lecture
Bernard aproveitou para dar uma volta /walk
Mas quem é que vai dar um subsidio/subsidize
Mas quem é que vai dar um subsidio/subsidise
a dizer-te que pode dar a impressão de/seem
into que, se puder, devo dar uma ajuda /help
into que, se puder, devo dar uma ajuda /aid
into que, se puder, devo dar uma ajuda /assist
ão de Persse, começa a dar exemplos/exemplify
: irracional impedia-a de dar voz /hear
: irracional impedia-a de dar voz /listen
e sentia-se capaz de dar um murro em/punch
ssel. Gostávamos de lhe dar uma palavrinha/speak
avrinha. -- Então vamos dar um aperto de/shake

EN verbs

religião, mas não se impo
religião, mas não se impo
religião, mas não se impo
à carreira de clínico geral.
à carreira de clínico geral.
à carreira de clínico geral.
à carreira de clínico geral.
me livrar da tropa -- expli
me livrar da tropa -- expli
pela cozinha. Ainda há-de
capaz para estudar os Oc
que não confio em ti
ao Council. Um deles abai
ao Council. Um deles abai
ao Council. Um deles abai
: O Espelho da Sinceridad
ao medo. Ainda não eram
ao medo. Ainda não eram
quem tentasse detê-lo. M
. -- Então vamos dar um a
mão. «Deixar que o seu

MACHINE TRANSLATION

Recognition of
Portuguese SVC
and translation
into English verbs

OpenLogos Resources at DFKI

- The Language Technology Lab of DFKI has adapted OpenLogos from the commercial Logos System
- OpenLogos is available in Linux platform using PostgreSQL
- OpenLogos employs only open source components
 - Use of open source development **tools and compilers**, such as GCC
 - Replacement of non-open **code and libraries**
 - Use of open **source databases** instead of a commercial database. All **language specific resources** have been converted to an open source database, such as PostgreSQL
 - Use of **open standards** instead of vendor specific protocols
 - As a proof of concept for the software migration, **Linux** is used as target **platform** for the first open source release of Logos

DFKI User Assistance with OpenLogos

DFKI hosts an open OpenLogos mailing list dedicated to discussion and exchange of information concerning OpenLogos developments and problems at <http://www.dfki.de/mailman/listinfo/openlogos-list>

Supplemental Free Resources at Logos System Archives Website

SAL representation language

- Described in depth

Detailed information about

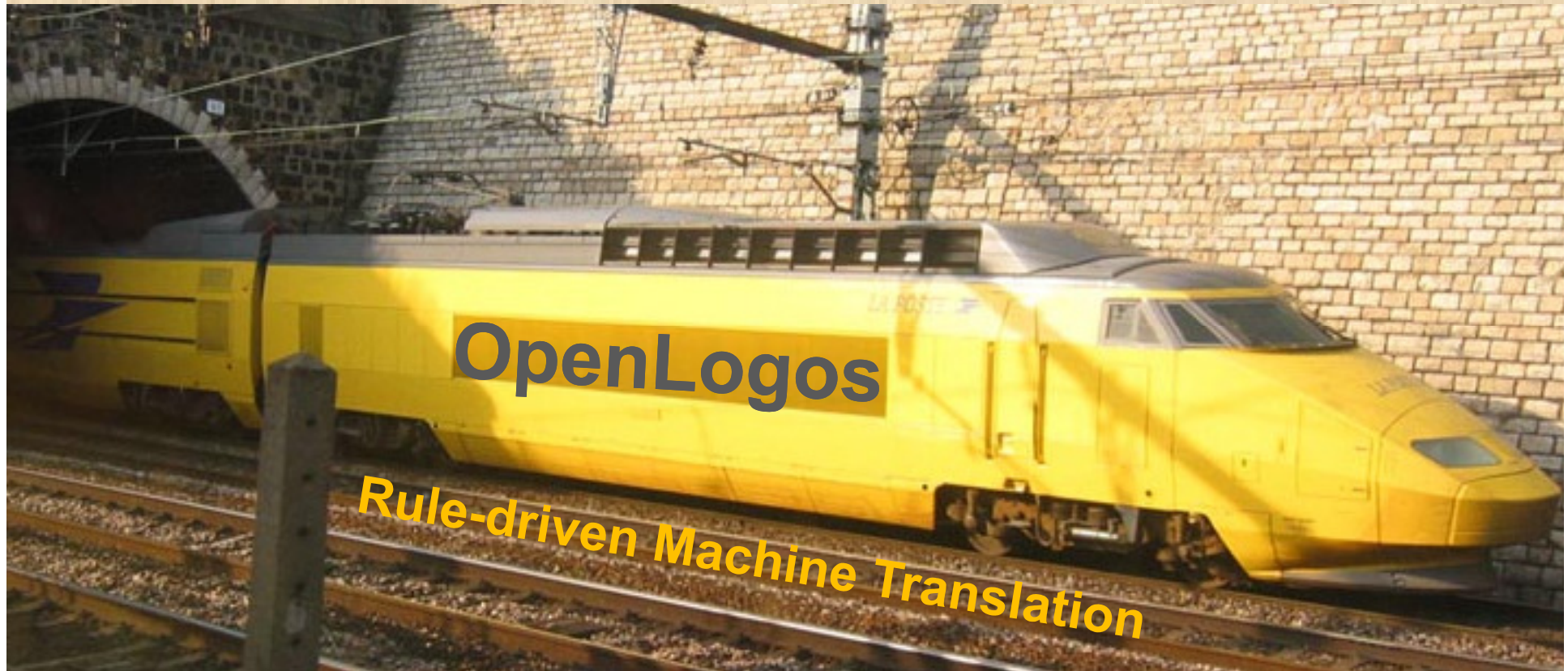
- System Architecture
- System Flow

Technical papers

On the Internet at: <http://logosystemarchives.homestead.com>

First International Workshop on Free/Open-Source Rule-Based Machine Translation

2nd and 3rd November 2009
Universitat d'Alacant, Alacant - Spain



Bernard Scott
logos.institute@gmail.com

Anabela Barreiro
barreiro_anabela@hotmail.com