An Open Service Framework For Next Generation Localisation

Stephen Curran
Trinity College, Dublin

CNGL

Centre for Next Generation Localisation

- Irish government funded research project into localisation
- TCD, DCU, UCD, UL
- Microsoft, Symantec, IBM, SDL, VistaTec

Research groups

- Language technologies
  - Machine translation, text analytics, speech synthesis and recognition
- Localisation processes and standards
- Systems framework
  - Software integration
  - Service Oriented Architecture for localisation

Presentation agenda

- What is localisation?
- Current localisation process and supporting software
- Problems with this software
- An open service-oriented framework for localisation
- Future work

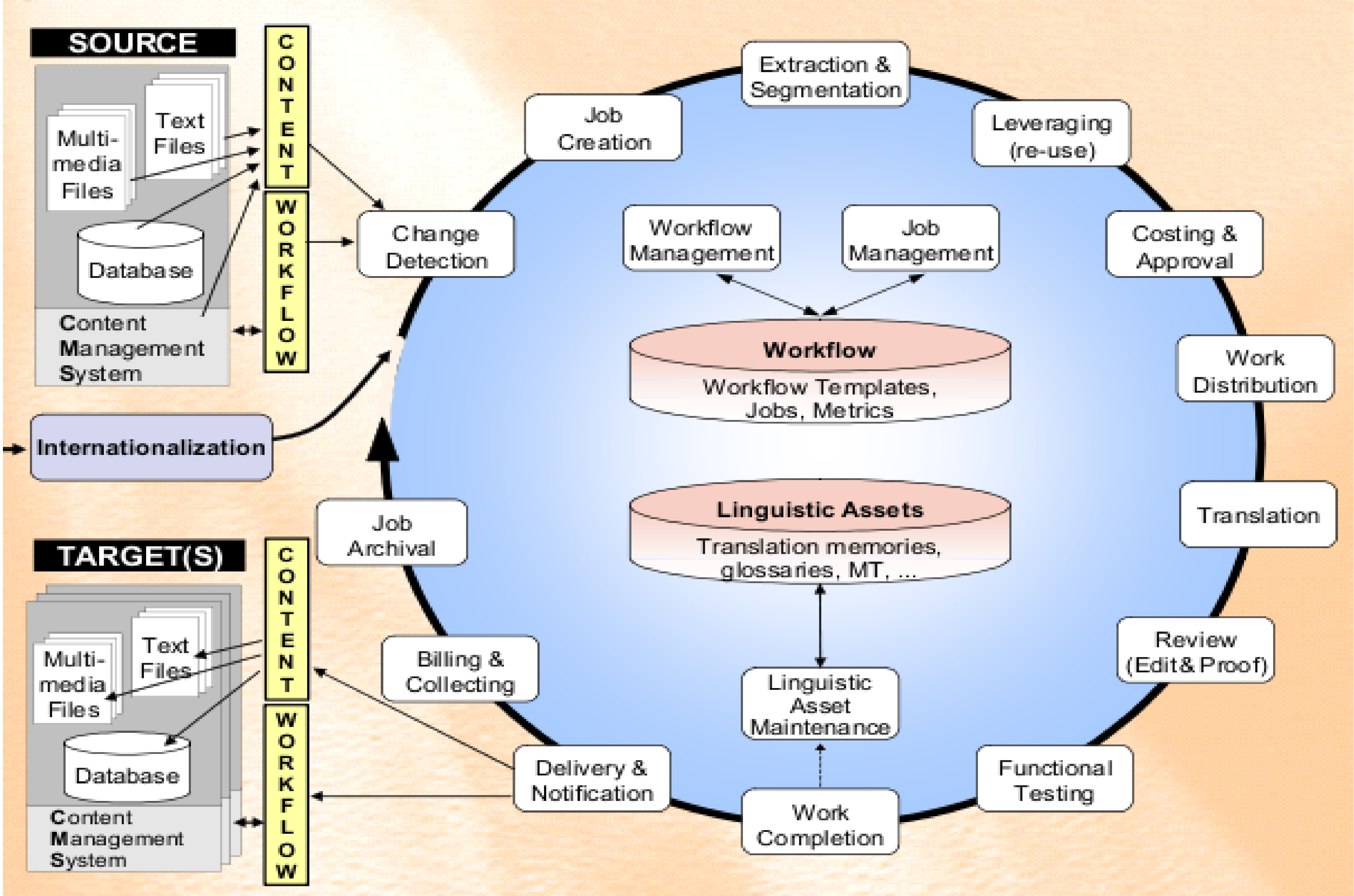
What is localisation

- Adaption of content to a locale or language
- Product localisation
    - Adaption of a product to sell into another language market
- Not just translation
    - Engineering work
    - Adaption of layout
    - Testing

Industrial partners

- Software localisation
- Content types
  - User interface
  - Product help
  - Printed documentation
  - Marketing content
  - Online user support pages

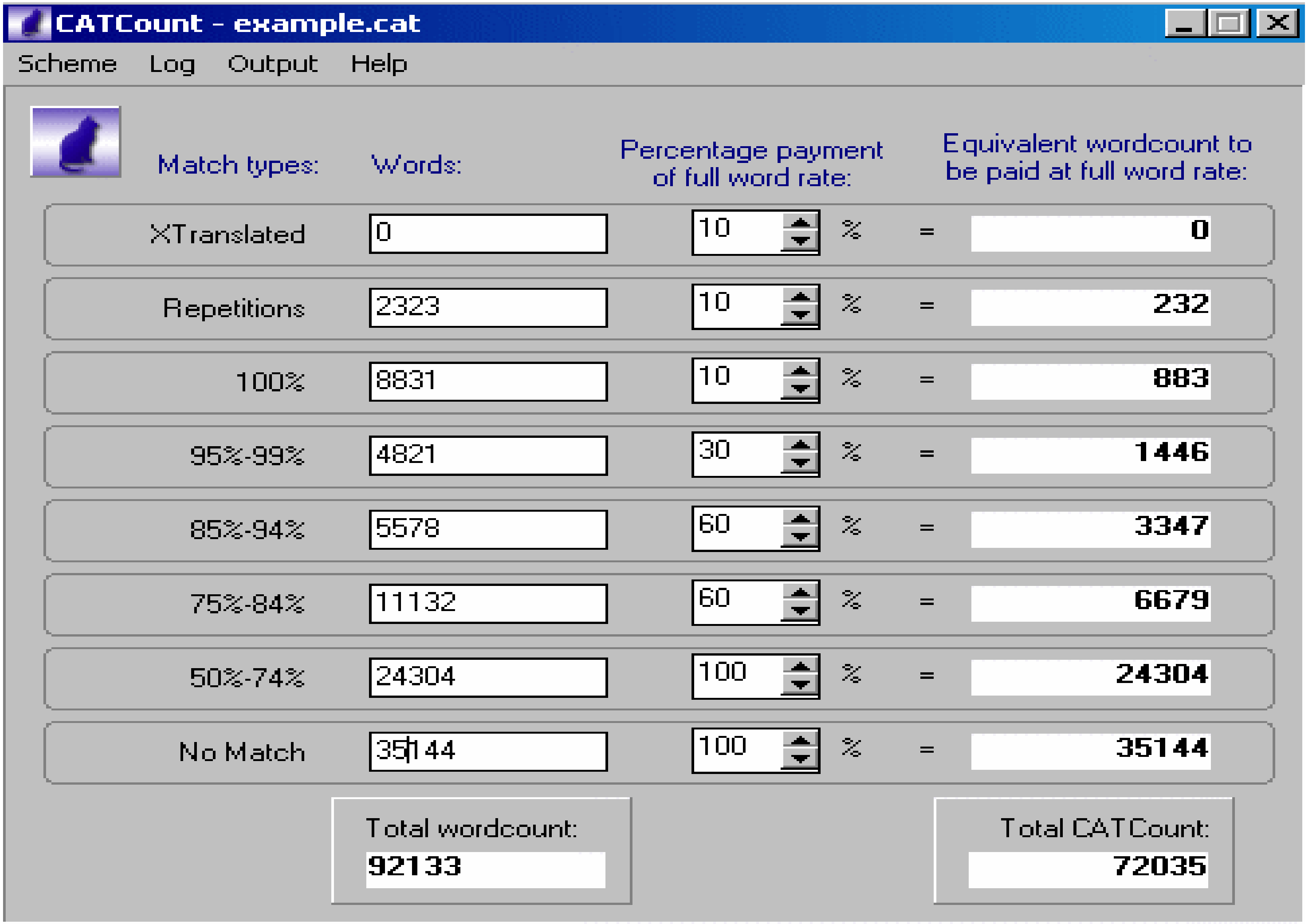
Localisation process



Translation memory

- Database of previous translations
- No need to translate the same sentence twice
- Exact match/fuzzy match
- Translation cost calculated based on translation memory hits

Translation memory leverage



Match types:	Words:	Percentage payment of full word rate:			Equivalent wordcount to be paid at full word rate:
XTranslated	0	10	%	=	0
Repetitions	2323	10	%	=	232
100%	8831	10	%	=	883
95%-99%	4821	30	%	=	1446
85%-94%	5578	60	%	=	3347
75%-84%	11132	60	%	=	6679
50%-74%	24304	100	%	=	24304
No Match	35144	100	%	=	35144

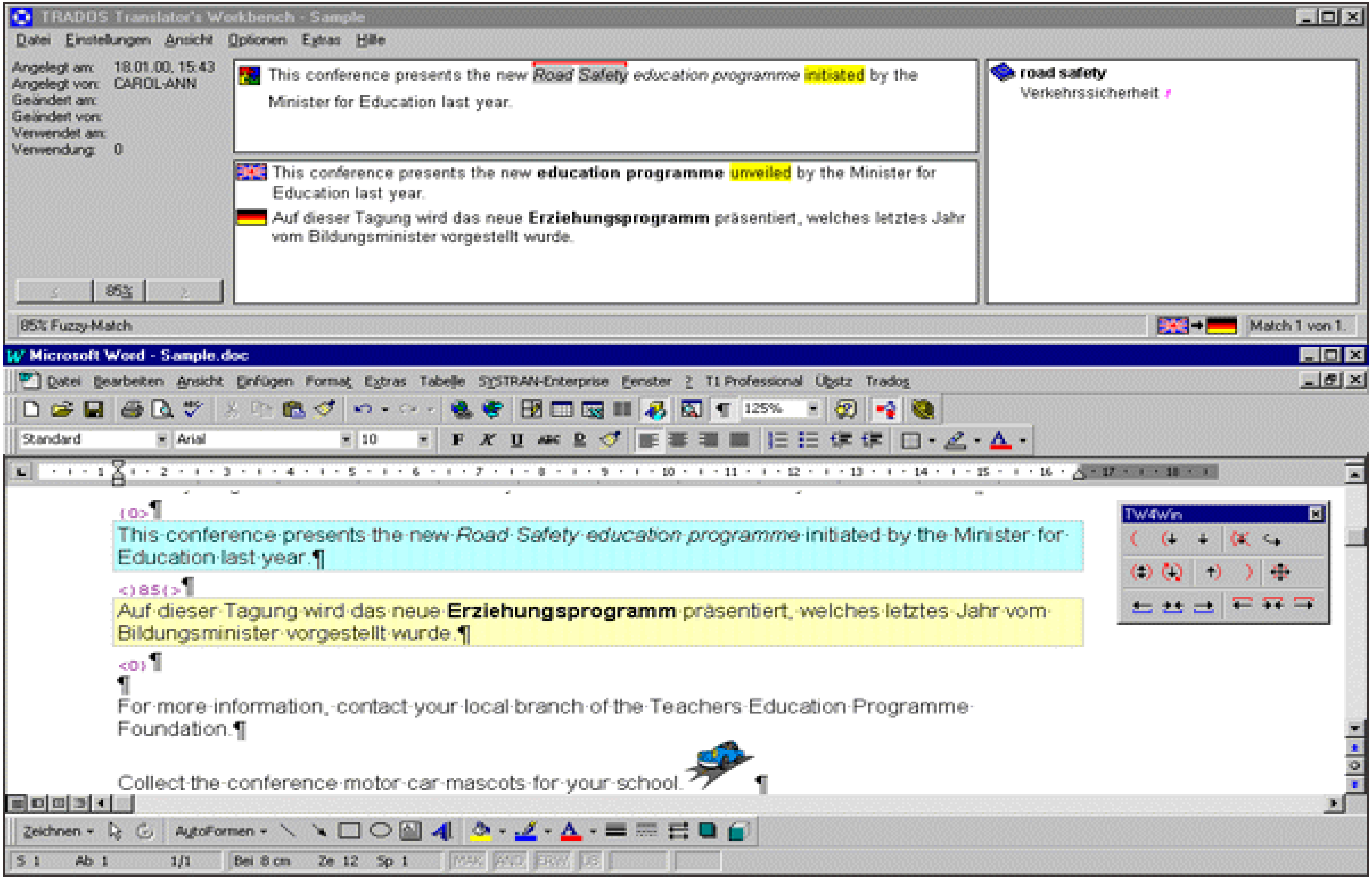
Total wordcount: **92133**

Total CATCount: **72035**

Termbase

- Translation of core terminology or phrases
- Ensures consistent translation of terminology across the product
- Definition of term to help translator

Computer assisted translation tool



Angelegt am: 18.01.00, 15:43
Angelegt von: CAROL-ANN
Geändert am:
Geändert von:
Verwendet am:
Verwendung: 0

This conference presents the new *Road Safety education programme* initiated by the Minister for Education last year.

This conference presents the new **education programme** unveiled by the Minister for Education last year.

Auf dieser Tagung wird das neue **Erziehungsprogramm** präsentiert, welches letztes Jahr vom Bildungsminister vorgestellt wurde.

road safety
Verkehrssicherheit

85% Fuzzy-Match

Match 1 von 1

<0>¶

This conference presents the new *Road Safety education programme* initiated by the Minister for Education last year.¶

<}85{>¶

Auf dieser Tagung wird das neue **Erziehungsprogramm** präsentiert, welches letztes Jahr vom Bildungsminister vorgestellt wurde.¶

<0>¶

¶

For more information, contact your local branch of the Teachers Education Programme Foundation.¶

Collect the conference motor car mascots for your school. ¶

TRADOS Translator's Workbench - Sample
Datei Einstellungen Ansicht Optionen Extras Hilfe
85%
Microsoft Word - Sample.doc
Datei Bearbeiten Ansicht Einfügen Format Extras Tabelle SYSTRAN-Enterprise Fenster ? T1 Professional Übstz Trados
125%
Standard
Arial
10
TW4Win
Zeichnen AutoFormen
S 1 Ab 1 1/1 Bei 8 cm Ze 12 Sp 1 MAK ÄND ERW ÜB

Translation management system

- Manages end-to-end process
- Management of :
  - Workflows
  - Users
  - Projects
  - Translation memories and glossaries

Localisation standards

- LISA standards
  - Translation Memory eXchange (TMX)
  - Termbase eXchange (TBX)
  - Segmentation Rules eXchange (SRX)
- OASIS standards
  - Localisation Interchange File Format (XLIFF)

Problems with current software

- Lack of support for data standards
  - Proprietary formats
  - Import but not export
- Lack of standard protocols/interfaces for common tasks
- Inflexible workflows
  - Support for a limited set of activities

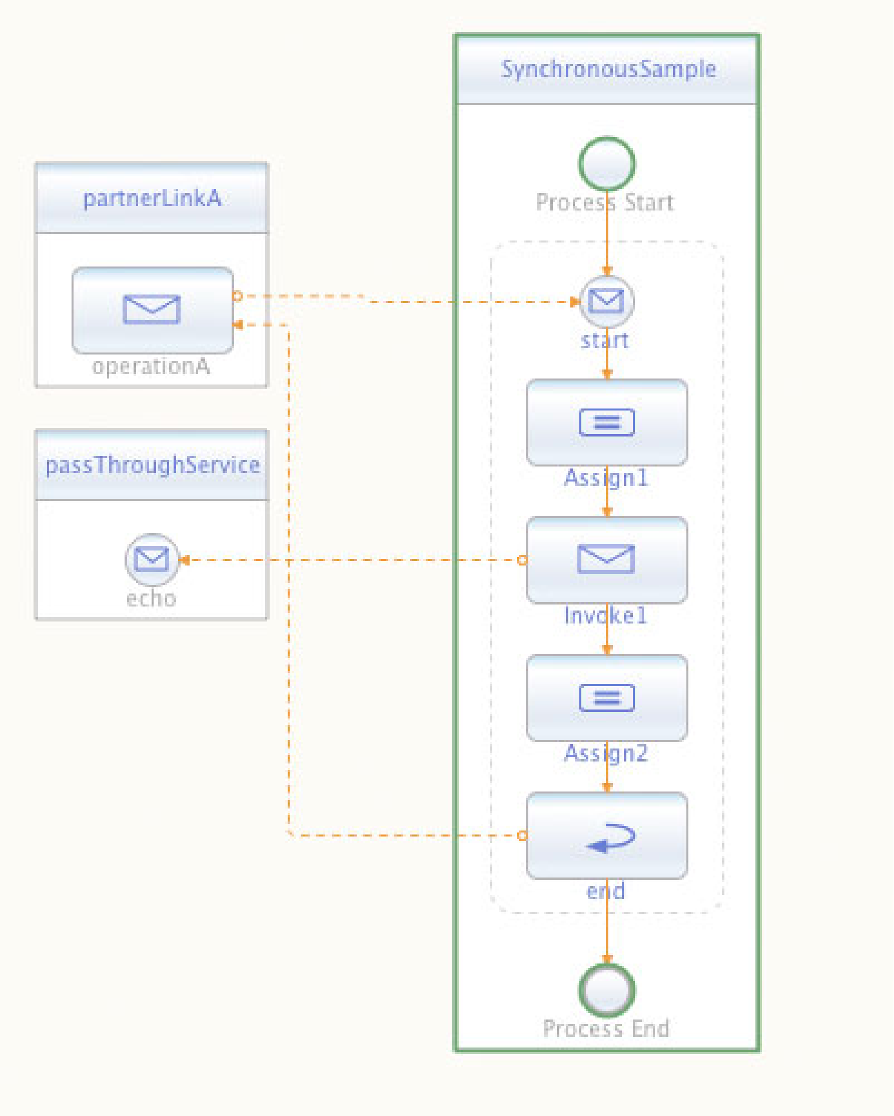
New service-oriented approach

- Support for localisation data standards
- Definition of interfaces for common localisation tasks
- Use of open web based protocols and technologies
    - HTTP
    - WSDL and SOAP
    - REST

Initial set of services

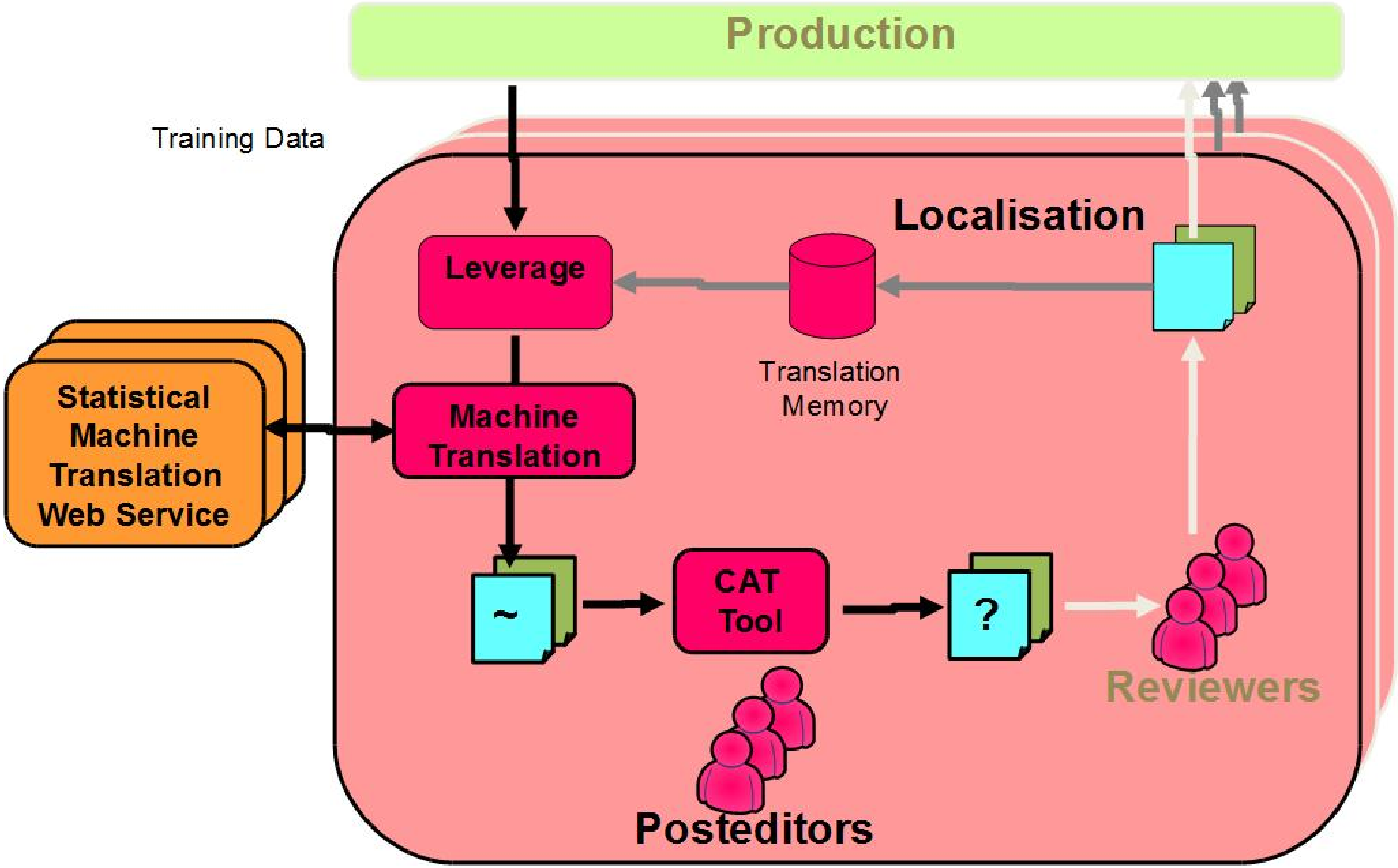
- Machine translation
  - Service wrapper for the Moses decoder
- Language classification service
  - Service wrapper for the TextCat library
- Domain classification service
  - Service wrapper for text classification algorithm from academic partner

BPEL

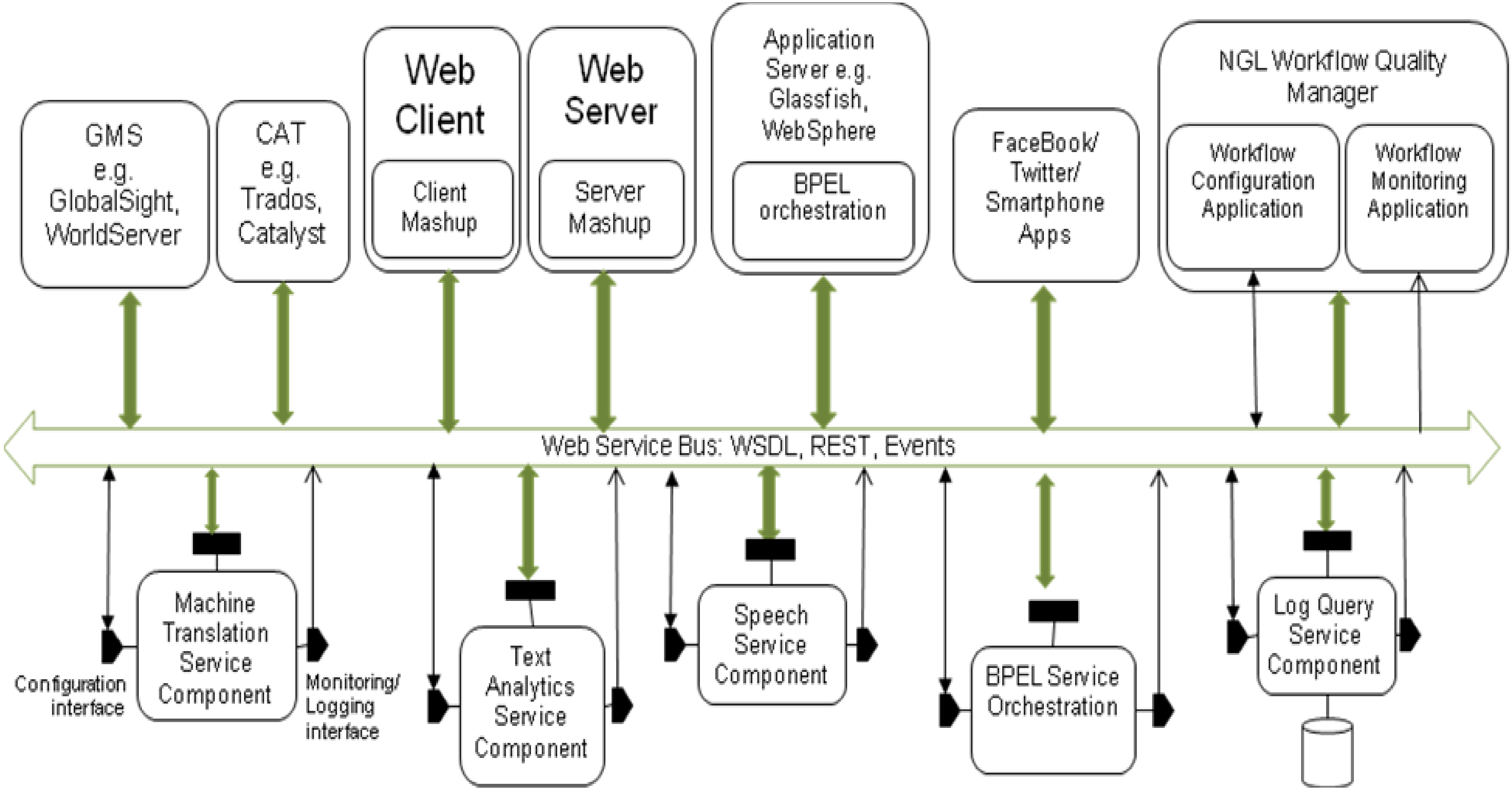


- Compose services into an executable workflow using BPEL
- XML based language
- Drag and drop designers available
- Netbeans and Glassfish

Plug into existing workflows



Service oriented architecture


GMS e.g. GlobalSight, WorldServer
CAT e.g. Trados, Catalyst
Web Client
Client Mashup
Web Server
Server Mashup
Application Server e.g. Glassfish, WebSphere
BPEL orchestration
FaceBook/ Twitter/ Smartphone Apps
NGL Workflow Quality Manager
Workflow Configuration Application
Workflow Monitoring Application
Web Service Bus: WSDL, REST, Events
Configuration interface
Machine Translation Service Component
Monitoring/ Logging interface
Text Analytics Service Component
Speech Service Component
BPEL Service Orchestration
Log Query Service Component

Future work

- Propose interfaces for common localisation components
- Implement these interfaces wrapping software made available to the project
- Expand on example processes