



SEVENTH FRAMEWORK PROGRAMME
THEME 3
Information and Communication Technologies

PANACEA Project

Grant Agreement no.: 248064

**Platform for Automatic, Normalized Annotation and
Cost-Effective Acquisition
of Language Resources for Human Language Technologies**

**D5.5: English-French and English-Greek
bilingual dictionaries for the Environment and
Labour Legislation domains**

Dissemination Level: Public
Delivery Date: June 30th 2012
Status – Version: Final v1.0
Author(s) and Affiliation: Yvette Graham (DCU), Gregor Thurmair (Linguattec),
Antonio Toral (DCU), Vera Aleksić (Linguattec)

Relevant Panacea Deliverables

D5.1 Parallel Technology Tools for Panacea
D5.4 Bilingual Dictionary Extraction Tools
D7.4 Panacea Deliverable D7.4: Evaluation Report (Third cycle)

Table of contents

Table of contents.....	2
1 Introduction.....	3
2 Languages and Domains.....	3
3 Process of Bilingual Dictionary Extraction	3
4 Conclusions.....	4
A. Example Dictionary Entries	5

1 Introduction

This report describes the deliverable D5.5 of the PANACEA project: Greek-English and French-English bilingual dictionaries for the environment and legislative domains automatically extracted from parallel corpora. This deliverable is one of the outcomes of the task WP5.2. The resulting domain-specific dictionaries include a total of 80,000 bilingual entries for two language pairs: English–French and English–Greek and two domains: Environment and Labour Legislation.

2 Languages and Domains

The domain-specific bilingual dictionaries are delivered for English–French and English–Greek in the domains of Environment and Labour Legislation (see Table 1).

<i>language pair/domain</i>	<i>environment</i>	<i>labour legislation</i>
English—Greek	√	√
English—French	√	√

Table 1: Language pairs and domains of bilingual dictionaries provided by D5.5

3 Process of Bilingual Dictionary Extraction

The task WP5.2 involved the development of tools for automatically extracting bilingual dictionaries from parallel corpora. Deliverable D5.3 included the delivery of parallel corpora for the required language pairs and domains. The bilingual dictionaries produced were constructed by inputting the parallel corpora of D5.3 to the automatic dictionary extraction tools of D5.4.

The standard dictionary extraction tool was not developed for the Greek–English language pair and therefore was only applied to the parallel corpus for French-English but for both domains, resulting in small high-precision dictionaries for each domain.

The advanced tool was applied to the same data to produce a larger (but slightly less precise) bilingual dictionary for each domain for French–English and also for Greek–English for both domains. No entries were filtered out due to low feature scores as f-score results were best without filtering as reported in the deliverable report for D5.4. In summary, a total of 4 bilingual dictionaries were produced for French–English and two for Greek–English.

The following table contains some statistics about the data used to produce the dictionaries and the dictionaries themselves.

			Parallel Corpus (sent pairs)	Phrase Table Entries	Dictionary Entries
French-English	Legislative	Basic Tool	20,261	763,136	4,557
		Advanced tool	20,261	852,512	27,798
	Environment	Basic Tool	10,240	405,471	2,856
		Advanced tool	10,240	444,651	17,690
Greek-English	Legislative	Advanced tool	7,064	415,100	13,220
	Environment	Advanced tool	9,653	501,062	18,584

Table 2: Detailed statistics of the delivered domain-specific bilingual dictionaries

4 Conclusions

We have successfully acquired bilingual dictionaries for two language pairs, French–English and Greek–English, in two required domains, Environment and Labour Legislation, of sufficient size and quality. An evaluation of the quality of the dictionaries produced by the tools used here will be covered in detail in D7.4 (month 34).

A. Example Dictionary Entries

A.1 Example Dictionary Entries for French-English for the Environment Domain produced by the Standard Dictionary Extraction Tool

ratifier	Vb	ratify	Vb
fossile	Ad	fossil	Ad
élaboration	No	preparation	No
activité de gestion	No	management activity	No
valeur cible	No	target value	No
carte de risques	No	risk map	No
dirigeant	Ad	leader	No
marché communautaire	No	community market	No
population humaine	No	human population	No
renard	No	fox	No
banque mondiale	No	world bank	No
premier jour	No	first day	No
transaction	No	transaction	No
spécifique moyenne	No	average specific	No
entité	No	unit	No
inondation	No	flood	No
aide	No	aid	No
développement économique	No	economic development	No
pétrole	No	oil	No
international	Ad	international	Ad
option	No	option	No
banque	No	bank	No
implication	No	implication	No
information statistique	No	statistical information	No
substance prioritaire	No	priority substance	No
externe	Ad	external	Ad
climatisation	No	condition	Vb
vulnérable	Ad	vulnerable	Ad
engagement de réduction	No	reduction commitment	No
mentionner	Vb	note	Vb
domaine suivant	No	following area	No
incinération	No	incineration	No
conséquence économique	No	economic consequence	No
banque européenne de reconstruction	No	european bank for reconstruction	No
exemption	No	exemption	No
sûreté	No	safety	No
argile	No	clay	No
doubler	Vb	double	Vb
obstacle	No	obstacle	No
critère suivant	No	following criterion	No
forum	No	forum	No
géologique	Ad	geological	Ad
déchets solide	No	solid waste	No
système de climatisation	No	conditioning system	No

A.2 Example Dictionary Entries for the Greek-English for the Environment Domain produced by the Advanced Dictionary Extraction Tool

αγριογαρύφαλλο	No	wild carnation	No
ακραίων θερμοκρασιών	No	extreme temperatures	No
ακραίων φαινομένων	No	extreme events	No
ακριβής	Ad	precise	Ad
ακτιβιστής	No	activist	No
ακυρώνω	Vb	counteract	Vb
αλλαγή χρήσεων	No	use change	No
αλλεργικός	Ad	allergic	Ad
άμεσος	Ad	urgent	Ad
αμέτρητος	Ad	countless	Ad
αμμονησίδα	No	sandy islet	No
αναγέννηση	No	regeneration	No
ανακύκλωση	No	re-cycling	Vb
ανακυκλώσιμα υλικά	No	recyclables	No
ανανεώσιμων καυσίμων	No	renewable fuels	No
αναπτυξιακά έργα	No	development projects	No
αναπτύσσω	Vb	build	Vb
ανατολικό τμήμα	No	eastern part	No
αναφορά	No	report	Vb
αντιδρώ	Vb	react	Vb
αντικρουόμενος	Ad	conflicting	Ad
αντικρουόμενων παραγόντων	No	opposing factors	No
αξιόλογος	Ad	sizeable	Ad
αξιόπιστο δείκτη	No	good indicator	No
απαιτούμενη συμπεριφορά	No	requisite behaviour	No
αποδεικτικών στοιχείων	No	evidence	No
αποδεκτός τρόπος διάθεσης	No	acceptable disposal route	No
αποθήκευση παγετώνων	No	glacier storage	No
αποκατάσταση	No	restoration	No
απομόνωση από τις κοινότητες	No	isolation from the communities	No
απόφαση	No	decision	No
αρνητικές επιπτώσεις	No	negative impacts	No