

Knowledge Rich Contexts: a Useful Feature for Translators

Outline



- Introduction
- The CRISTAL Project
- Testing KRCs with Translators
- Experiment and Protocol
- Some Results
- Conclusion

Introduction: Corpora and Languages

Corpora have been used in different fields related to Languages:

- Translation Studies, Language studies (Toury 1995)
- Language learning (Aston 1999)
- Terminology (Pearson 1998) (Bourigault 2001)
- Translation training (Baker 1992), (Aston 1999), (Bernanini 2000)
- **by translators (Aston 2000), (Somers 1998)**

1. The CRISTAL Project and KRCs

1.1 The CRISTAL Project

1.2 KRCs

1.3 Objectives of the Project

- **CRISTAL = Contextes Riches en connaissances pour la TrAduction TerminoLogique**
- Funded by French ANR agency (2012-2015)
- 4 partners :
 - LINA lab., University of Nantes
 - CLLE-ERSS lab., University of Toulouse
 - FTI, University of Geneva
 - Lingua et Machina company

Knowledge Rich Contexts (KRCs)

- General definition
 - « By **knowledge-rich context**, we designate a context indicating at least one item of domain knowledge that could be useful for conceptual analysis » (Meyer, 2001: 281)
 - KRCs come from research in the area of Terminology Knowledge Base building, e.g. (Condamines 2001)
 - Taken up in Translation, e.g. (Bowker, 2001)

Knowledge Rich Contexts (KRCs)

- **Example of conceptual KRC**
 - *“**Tephra** is any material ejected explosively from a volcano (**ash**, **lapilli**, **cinder**, and **spatter**)”*
 - *In a KRC, terms (e.g. “Tephra” and “ash”) are related by a **conceptual** relation: here definition and hypernymy*
 - *A **marker** shows this relation, here: “is”*
(Barrière 2006) calls it a Knowledge Pattern
- **Example of linguistic KRC**
 - *“Entre deux plaques divergentes, se trouve une zone d'accrétion où **s'épanchent** des **laves** fluides et denses, de type basaltique.”*
 - *Here, terms are related by a relation of **collocation***
 - *Here: verb “s'épancher” and “laves” are collocates*

General objectives of the project

- Characterize KRCs as a linguistic object (Toulouse)
 - list of KRC markers
- Extract automatically KRCs from Comparable Corpora (Nantes)
 - algorithms that extract KRCs using markers and collocates from a list of terms
- Test KRC use with translators (Toulouse, Nantes and Geneva)
 - a prototype interface and experiments with translators

Some outputs

● A list of KRC markers

	French	English
hypernymy	33	35
meronymy (part-of)	98	99
cause	338	411

- Algorithms that extract lists of KRCs
 - For a given list of terms

Examples of Automatically extracted KRCs

éruption	Ce type d'<term>éruption</term> est la marque des "volcans rouges" dits effusifs, propres aux zones d'accrétion et de point chaud	Hypo	lave:gaz
éruption	Ce dégazage s'exprime sous la forme de spectaculaires <term>fontaines</term> de laves	COLLO-ADJ	spectaculaire
vesicle	Cinders are glassy particles that contain <term>vesicles</term>	HOLO	cinder
vesicle	Reticulite has a still lower density, with <term>vesicles</term> occupying up to 98% of the total volume	COLLO-ADJ	occupy

2. Testing KRCs with Translators

2.1 Previous Experiments about the Use of Translation Resources

2.2 Our Goal

Previous Experiments

- Experiments on the use of different resources by translators:
 - Varantola 1998: dictionaries
 - Bowker 1998: corpora
 - Künzli 2001: various resources
 - Desilets 2009: corpora
 - Delpech 2011: extracted biterms

Our goal

- Test the usefulness of KRCs for translation
 - KRCs are to be considered as a special kind of translation resource
- Use an interface similar to current CAT tools
- Test with both trainee and professional translators
- Get a sizeable number of participants
- Record / monitor the test

3. The Test Protocol

Participants and Tasks

- A pilot study in Geneva (Dec. 2013)
- March 2014 (focus presentation)
 - 18+24 = 42 Master2 students in Angers+Toulouse
 - Translate a text
 - EN => FR
 - 150 words
 - Popular science text (volcanology)
 - Validate KRCs seen as useful for translation (source and target languages)
 - A form to be filled after the translation
 - Oral discussion (group or individual on video)
- April 2015: trainee and professional translators

continuer, merci de choisir les meilleurs textes.

Basalt cinder cones are the most common kind of volcano worldwide. They are also some of the smallest volcanoes. A typical eruption goes through two stages. The first is called the

Les cônes de cendres

Source Cible

CRC	
1	Cinder cones are conical hills produced by the accumulation of cinders and other tephra.
2	"A cinder cone is a "conical hill formed by the accumulation of cinders and other pyroclasts, normally of basaltic and andesitic composition" (Bates and Jackson 1980, p. 112). "

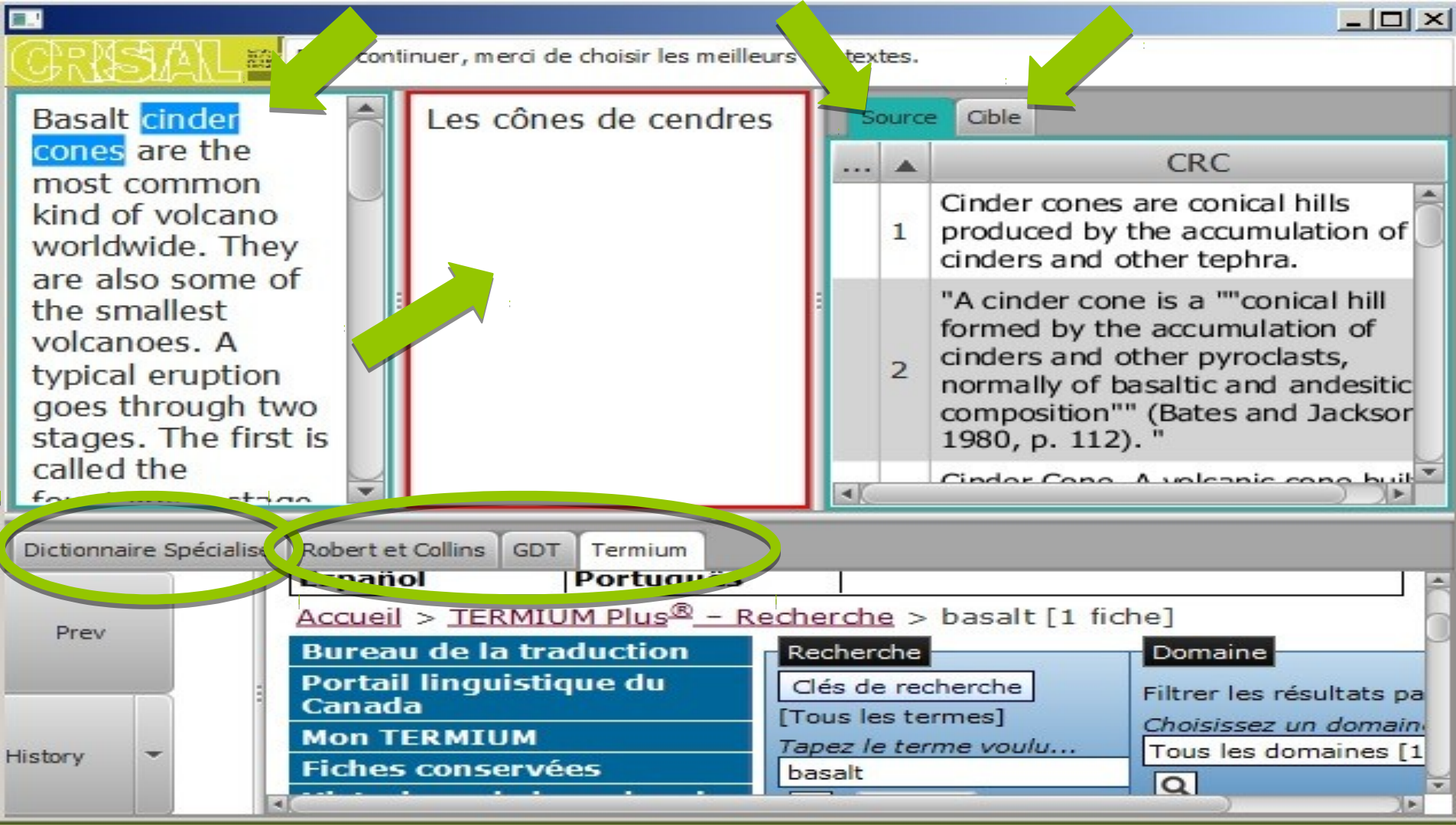
Dictionnaire Spécialisé Robert et Collins GDT Termium

Accueil > [TERMIUM Plus®](#) - Recherche > basalt [1 fiche]

Bureau de la traduction
Portail linguistique du Canada
Mon TERMIUM
Fiches conservées

Recherche
Clés de recherche
[Tous les termes]
Tapez le terme voulu...
basalt

Domaine
Filtrer les résultats par
Choisissez un domaine
Tous les domaines [1]



Monitoring

- **While translating**
 - Argos logs :
 - Keyboard activity: written text, shortcuts
 - KRCs : searched term + KRC displayed+validated
 - Other resources looked at
 - Window shifting
 - BB Flashback records screen activity and keyboard
- **Post-translation**
 - Interviews recorded (Audacity, then transcribed)

Results interpretation

- We used Triangulation (ex. Alves, 2003, Denzin 1978)
 - “cross-checking the results one set of data provides with results from another set of data” (Saldanha & O’Brien, 2013 : 23)
 - Online questionnaire
 - Logs
 - Argos log
 - Recorded Screen Video
 - Post-translation interviews
 - Group interviews, semi-directed
 - Individual self video replayed-based interviews

3 Translators faced with KRCs: some results

3.0 Some figures

3.1 Preliminary remarks: a few hindrances

3.2 Defining what a useful KRC for translators is

3.3 Resource complementarity

3.4 Interface ergonomics

3. Some figures

- We suggested poor and rich contexts
Most of the selected contexts were rich ones

	KRC	Conceptual	Linguistic
Source language			
20 most selected contexts	90%	65%	25%
Target language			
20 most selected contexts	90%	55%	35%

Some figures

- KRCs looked at for "cinder cone"
- KRC 1, 2 et 10 are preferred (by 11,13 and 12 testers)

id															S
1	1	1	1		1	1	1	1		1	1	1	1	11	
2	1			1			1	1	1	1	1	1	1	13	
3				1	1			1	1				1	5	
4	1				1	1			1	1			1	6	
6				1						1		1	3		
10	1	1			1	1	1	1	1		1		1	12	

- Rich(def.): 10. **Cinder cones** are among the most common volcanic landforms found in the world.
- Poor: This photo shows the interior of Strawberry Knoll, a **cinder cone** in southwestern Utah

3 Translators faced with KRCs: some results

3.0 Some figures

3.1 Preliminary remarks: a few hindrances

3.2 Defining what a useful KRC for translators is

3.3 Resource complementarity

3.4 Interface ergonomics

A few hindrances

- 1) “ecological validity” (Ehrensberger & Massey, 2008)
- Situation similar to real translator’s work, yet, not exactly the same
 - KRC Validation was compulsory

Were you bothered by the compulsory validation of KRCs during your translation

YES

69.05%

- 2) technical bugs

3 Translators faced with KRCs : some results

3.0 Some figures

3.1 Preliminary remarks: parasitic elements

3.2 Defining what a useful KRC for translators is

3.3 Resources complementarity

3.4 Interface ergonomomy

KRC for translators according to the *literature*?

- Bowker (2011, 2012) & Varantola (1998)
 - Information about usage (collocations)
 - Info. about term frequencies
 - Info. about conceptual relationships between terms (synonymy, meronymy, hypernymy)
 - Pragmatics (style, register, genre)
 - Info. about usages to be avoided
 - « any context that contains useful information » (Bowker 2012)

A useful information can have a broad coverage

- « Translators might not even know what they need: they are seeking inspiration, associations, similar examples, parallel situations that can be adapted »

Bowker 2012

« it is often a case of I don't know what I'm looking for, but I'll recognize it when I see it »

A useful KRC according to our *experiments*?

- A useful KRC is a context that:
 - 1. looks like the source segment to be translated
 - 2. comes from a trusted source
 - 3. can apply to more than one term
 - 4. is neither too short nor too long
 - 5. Where the searched term is highlighted
 - *Hint* : 6. holds information that can be used at different steps of the translation process

- 1. looks like the source segment to be translated

- Term : cinder cone
- The following KRC chosen by **12/42** testers
 - *N° 10. **Cinder cones** are among the most common volcanic landforms found in the world.*
- Text to be translated :
 - *Basalt **cinder cones** are the most common kind of volcano worldwide.*

□ *Hint* : 6. That contains a lot of information

- ETU2: What's more, a lot of information was provided, sometimes for a single term, I really needed that context (...)

3 Translators faced with KRCs : some results

3.0 Some figures

3.1 Preliminary remarks: parasitic elements

3.2 Defining what is a useful KRC for translators

3.3 Resource complementarity

3.4 Interface ergonomics

Resource complementarity

Contexts and lexicographic resources are complementary

Agrees and partially agrees

85.74%

I could have worked without contexts

Does not agree and does not partially agree

64.29%

I could have worked only with contexts, without classic specialized resources

Does not agree and does not partially agree

78.57%

Resource complementarity

- Complementarity about KRC types:

- Source KRCs/target KRCs - Conceptual/linguistic KRC
 - Source KRC: comprehension, definition
 - STUD1: As for me anyway, **in the source contexts, having a definition was of help**, it really helps to understand what is all about.
 - Target KRC: collocations, verifications
 - STUD1: And in source and target contexts, **you can find definitions and especially collocations**, and which goes with which verb, that kind of things, so you can really render the appropriate style as well as possible.

Resource complementarity

- Complementarity between resources types
 - 2 attitudes, according to the translation stage:
 - Starting with KRCs
 - Documentation, preparation
 - Starting with lexicographical resources
 - Bilingual resources allow transfer to turn to target KRCs
 - KRCs help choose between different lexicographical resources

Resource complementarity

		<u>sKrc</u>	<u>sKrc</u> <u>val</u>	<u>tKrc</u>	<u>tKrc</u> <u>val</u>	<u>spec</u> <u>dict</u>	<u>R&C</u>	<u>Term</u>	<u>GDT</u>	<u>Trans</u>
		1	2	3	4	5	6	7	8	9
action n°	1	19	0	1	0	5	6	10	1	0
action n°	2	1	18	0	1	8	4	7	1	2
action n°	3	2	8	2	0	12	4	10	2	2
action n°	4	5	4	2	3	10	4	5	5	4
action n°	5	0	8	3	2	10	2	4	4	6
action n°	6	0	8	3	2	10	1	4	4	10
action n°	7	0	3	4	5	7	0	6	3	14
action n°	8	1	3	5	5	5	0	5	0	18
action n°	9	3	4	2	8	2	3	4	2	14
action n°	10	0	5	3	8	8	2	3	2	11
action n°	11	0	2	3	6	12	0	3	0	16

Different strategies while translating :

- 19 students start looking at source KRCs, then 18 validate them
- fortunately the vast majority do not start by translating! (last column)
- target KRCs are used afterwards

Resource complementarity

- **back and forth:**
 - ETU1(VIDEO-based debrief): I tried to look up <D: Ok> when... I thought I would check in several... err <D: Oh, if you chose one> yes I did <D: Ok> and then I thought I... <D: you looked it up again > I looked up Termium <D: Ok>, Yes <D: that's it> yes, in fact I shifted from one to another (...) I like to check other resources, because <D: Yes right > not to...

Resource complementarity

- The interface has to be built so as facilitate the use of this complementarity:
 - Hence the importance of an interface that allows access to all information at once

3 Translators faced with KRCs : some results

3.0 Some figures

3.1 Preliminary remarks: parasitic elements

3.2 Defining what is a useful KRC for translators

3.3 Resources complementarity

3.4 Interface ergonomics

Interface ergonomics

- Features the testers liked: simplicity, short-cuts, resizable windows, easy to catch-up, all information in a single window

Easy to understand and to use

Online form

The displayed interface is **clear and easy** to use because it contains the essential elements only, you don't get lost in extensive menus, **and so you gain in efficiency**.

Interface ergonomics

Did you think that the interface contains too many windows to handle it easily?

NO	95.24%
----	--------

Did you appreciate working with an interface that gathers all available resources?

YES	90.48%
-----	--------

Did you appreciate being able to customize your interface (window size)?

YES	97.62%
-----	--------

Did you find the integrated notepad useful?

YES	47.62%
-----	--------

Conclusion: context richness: a complex notion

4. Discussion

- The notion of KRC for translators is a difficult and relative one, at least in terms of:
 - the translation stage
 - translator's habits
 - the type of other resources also used/usable

4. Discussion

- Translator's habits
 - ETU1: I should have asked before, but I did not understand what it was about, and, well, what we do is to **work with resources we are used to**, Grand Dictionnaire Terminologique, Termium, ...

4. Discussion

- The translation stage
 - Comprehension
 - Conceptual KRC
 - Source and target KRCs
 - Transfer
 - Target KRC
 - Conceptual and linguistic KRCs
 - Revision
 - Mainly linguistic KRCs

4. Discussion

- Types of resources used/usable
 - Complementarity

Online form

I would have appreciated being able to look for **pictures** (for the size of a softball).

➤ Source Text:

- During their flight through the air, the blobs cool to solid pieces called cinders (also called *scoria*), **which are typically no bigger than a softball.**

Merci!

josselin@univ-tlse2.fr

emmanuel.planas@univ-nantes.fr

aurelie.picton@unige.ch