

System for Translation Information in General (STIG) for everyone

Diana Santos

d.s.m.santos@ilos.uio.no

UiO, 31 October 2017



The past

Once upon a time by being introduced to the “Female Robinsonades bibliography”, I came up with the idea that a similar system for storing translation information should be advantageous for anyone dealing with translation studies – and a lot of people seemed to be doing this in ILOS, so I could suggest something for the good of all ILOS.

- To my surprise, the proposal was accepted, and proposed by DMLF to be implemented in the fall 2015.
- I did a brainstorm at ILOS at 28 September 2015, although only Hildegunn and Bergljot came, see <http://www.linguateca.pt/Diana/download/AprReqPANTERA.pdf>
- The first meeting with DMLF, however, happened 30 November 2015
- We started working in December 2015

As anyone who is involved in software projects knows, what takes longest is establishing real communication. And here we are!



What was to be STIG

A system to organize information about translation, above basic reference to data included in a parallel corpus.

- The name is to pay honour to Stig Johansson, who was a pioneer of parallel corpus research worldwide, and who inspired most parallel corpora after him (ENPC)
- The idea was to create a flexible enough system that could encompass all information that might make sense to store about a particular translation
- The original idea was to be able to have several views of the material in PANTERA, by time, place, translation time, etc.
- And then all sorts of interesting searches and/or aggregations started to develop: books which had more than 3 reviews; texts that had been translated more than four times; translators which had translated the same author...

What is STIG now

- STIG surpassed itself, in that it is now a generic system using topic maps.
- So several of its new instances do not deal with translation at all!
- But for the current presentation we'll stick to the original (and hard enough) problem...

Why is the usual way not enough?

Imagine that you have a... spreadsheet.

- With information about the specific pair of text excerpts you selected for your corpus
- Author, title, publisher, publication date, translator, page numbers.

This is fine whenever you want to know what was the source of the corpus contents.

BUT... what if you want to know more and get to visualize the entire field?

If you want to know more...

You cannot associate to a bibliographical reference the following:

- When was the first edition (of the translation) published? By which publisher?
- And about the original work? When was it written? Which version was used for the translation?
- Are there reviews of that translation, of that original work, of that collection which included it?
- Has the author been translated before (into the same language or other)? Has he received any prize?
- What other translations were used by the translator? From which languages?
- Has there been more translations after the piece was included in an anthology – in the source language, or in another language?
- Biographical details of the author and her relationship to other actors (translators, publishers, reviewers)

This is how STIG was born!

A database is not enough

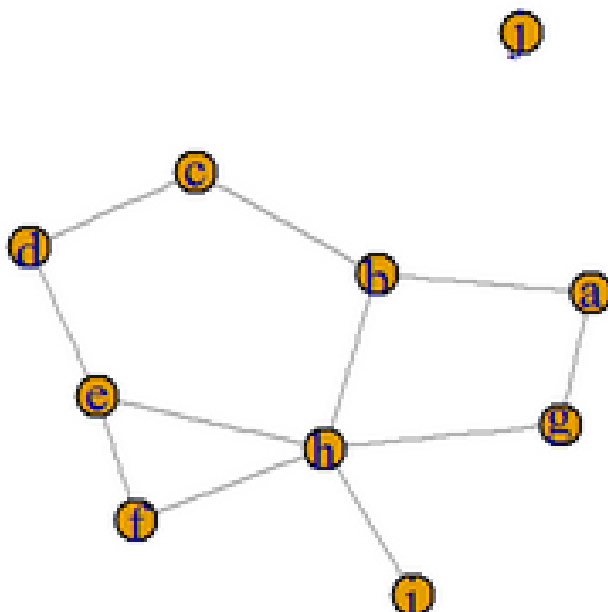
A database is a form of organizing efficiently information about one kind of data, around a key. Or several (few) keys. It does not allow easily associations between different kinds of entities. They should be values under the primary key.

But you do not want information about text excerpts only, but about all sorts of entities: texts as abstract objects, concrete publications of those texts, translators, authors, translations as abstract objects, publishers, literary critics, reviewers, literary prizes...

So you need to have a topic map system, that works with different entities, each with different properties, none of which being more important than the others, and which can be freely associated and have multiple associations.

Graphically

You can pick any entity and search from it, you are not forced to follow a rigid structure defined when you organized your data.



How to define your data model

It is complicated. It is like designing as many databases as entities you want to have in your system.

- Which properties does each entity have?
- How many properties do they share with other entities?
- Does it make sense to have similar values around?
- Are some entities just a specialization of others? In that case, does it make sense to repeat everything?

The example of the capital: is this a property of a city? Or an entity in itself?

How to define your data model: What is an author?

STIG
System for Translation Information in General

Search

Admin

Entity Types: **Advanced Search**

Autor

Nome

- Accioly, Breno 2
- Agualusa, José Eduardo 2
- Alcoforado, Mariana 2
- Almada Negreiros, José Sobral de 1
- Almada Negreiros, José de 1
- Almeida, Germano 2
- Amado, Jorge 2
- Andrade, Mário de 2
- Andresen, Sophia de Mello Breyner 2
- Antunes, António Lobo 2
- Askildsen, Kjell 2
- Athayde, Celso e MV Bill 2
- Athayde, Roberto 2
- Azevedo, Francisco 2

Autor

- Accioly, Breno ()
- Accioly, Breno (), brasileiro
- Agualusa, José Eduardo ()
- Agualusa, José Eduardo (), angolano
- Alcoforado, Mariana ()
- Alcoforado, Mariana (), português
- Almada Negreiros, José de (), português
- Almada Negreiros, José Sobral de ()
- Almeida, Germano ()
- Almeida, Germano (), cabo-verdiano
- Amado, Jorge ()
- Amado, Jorge (), brasileiro
- Andrade, Mário de ()
- Andrade, Mário de (), brasileiro
- Andresen, Sophia de Mello Breyner ()
- Andresen, Sophia de Mello Breyner (), portuguesa
- Antunes, António Lobo ()
- Antunes, António Lobo (), português
- Askildsen, Kjell ()
- Askildsen, Kjell ()

What is an author?

- Is an author anyone/anything that appear as author in a book cover or a ISBN register?
- Can an author have several designations and still be the same author?
- Are (pt.) autores and (no.) forfattere different?
- Can one be at the same time an autor and a (different) fortatter?
- Who decides?

Who decides? The information engineer / the information architect. Us, users who put the (agnostic) system to use.

NB! The most important thing is to document the decisions and their rationale, when it exists.

The knowledge engineer decides...

- Which entities types
- Which properties
- Which kind of values
- Which associations
- How to display them

This is an error-prone process, but there's no way of escaping it.

Is it worth while?

- One advantage is to understand the “world” better,
- therefore organizing the knowledge consistently,
- which allows one to retrieve it more accurately by making use of advanced search.

There is always a tradeoff between specialization and broad applicability. There is never a unique answer to any not-simple question. See http://www.asimovonline.com/asimov_FAQ.html for an interesting example.

Concluding...

- After demonstration of current capabilities...
- How do new users take possession of STIG for their language-pairs, or for their specific domains?

You can read *STIG's User manual and documentation* (2nd version) as a detailed description of what I had to do to apply it to PANTERA. But it shows also that it is a lot of work, because STIG is an extremely powerful system.