

# FastKwic, an “intelligent” concordancer using FASTR

Veronika Lux-Pogodalla<sup>1,2</sup>   Dominique Besagni<sup>1</sup>   Karën Fort<sup>1</sup>

<sup>1</sup>INIST-CNRS, 2 allée de Brabois, 54500 Vandoeuvre-lès-nancy

<sup>2</sup>ATILF-CNRS, 44 avenue de la Libération, 54000 Nancy

May, 2010



- 1 Outline
- 2 Introduction
- 3 FastKwic
- 4 Implementation in the TermSciences website
- 5 Conclusion

# Contexte

- TermSciences: more than 540,000 terms BUT no definitions.
- Large collection of bibliographical records at INIST.
- Previous work on term variation [Jacquemin 1994, Jacquemin and Royauté 1994, Jacquemin 1997, Royauté 1999].

⇒ Wish for a concordancer.

# Elements for the specification

- No complex request language.
- Mono- and multi-word terms ...
- ... occurring in texts with several variations.

# Term variations (Jacquemin, Royauté)

## Example

- *gamma-linolenic acid / gamma linolenic acid /  $\gamma$ -Linolenic acid.*
- *stuctural gene / structural erm gene.*
- *structural gene / structural and regulatory genes.*
- *resistance mechanism / mechanism of claritomycin resistance.*

## Major types

- typographical variations (ex. with or without -).
- morphological variations (ex. plural).
- syntactic variations (ex. insertion, coordination, permutation).

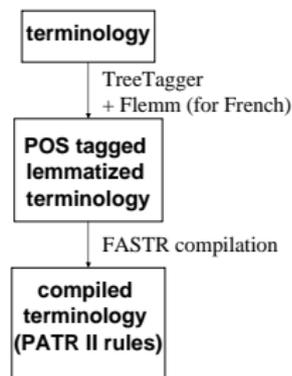
# Package features

- Two UTF-8 compliant Perl modules.
- Depends on several external freely available tools/resources
  - ▶ FASTR [Jacquemin 1997, Jacquemin et al. 1997],
  - ▶ TreeTagger [Schmid 1997],
  - ▶ Flemm [Namer 2000]
- Freely available on <http://www.cnrtl.fr/outils/>.

# Usage

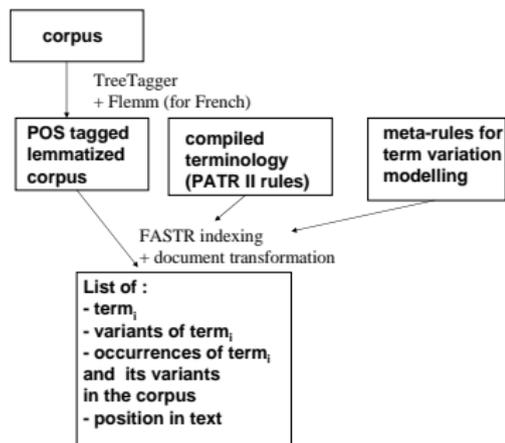
# Usage

- Resource compilation.



# Usage

- Resource compilation.
- **Indexing.**



# Usage

- Resource compilation.
- Indexing.
- Production of a concordancer.

```

<?xml version='1.0' encoding='UTF-8'?>
<Concordance>
  <Term>
    <TotalNumber>2</TotalNumber>
    <Preferential>
      <String>Gene amplification </String>
      <Number>2</Number>
      <Occurrences>
        <Occurrence>
          <Reference>000007</Reference>
          <Position>1:32</Position>
          <Transform>XX,25,Perm</Transform>
          <Context><b>Amplification of the MYC gene is</b> associated with
dmi</Context>
        </Occurrence>
        <Occurrence>
          <Reference>000008</Reference>
          <Position>1:38</Position>
          <Transform>XX,15,Ins</Transform>
          <Context><b>This gene facilitated amplification of</b> a 407-bp DNA
fragme</Context>
        </Occurrence>
      </Occurrences>
    </Preferential>
  </Term>

```

# Limitations

- Use of TreeTagger with limited context  $\Rightarrow$  errors in the POS tagging.
- Particular linguistic entities (*1,3,4-thiadiazole(2-amino)*) and their variants not taken into account.

## Resources, corpus

- Terminological resource = more than 540,000 terms ⇒
  - ▶ FASTR: only for terms with standard linguistic patterns.
  - ▶ IRC3(Royauté): for geographical names, drug names, chemical compounds, etc.
- Corpus = 30,744 records for French, 398,952 for English.
- Result of indexing put in MySQL database and accessed from <http://www.termsciences.fr/-/Index/Search/Concordancer/>

# Result example


**TermSciences**  
 Portail terminologique multidisciplinaire

Sign in | Help |  Français

[Consult](#) [TermSciences ?](#) [Link directory](#) [Services](#)

Concorderancier

DOCUMENTARY ZONE

Term  language fra

Enter term. The term will be automatically completed according to the contents of the terminological database

**Terminological zone**

1. Check terms in the left part.
2. Search for terms and their variants into the PASCAL 2005 concordancer.

[Concorderancier](#) [Web](#) [INIST](#)

Display results for domain :

Display results for variant :

**8 results of about 27745370**

1. de l'intérêt du conte pour le **développement affectif** et social d'enfants aveugles
2. est son lien intense avec le **développement affectif** et les apprentissages. Les au
3. du foetus, mais aussi sur le **développement psychomoteur et affectif** de l'enfant. Cet article res
4. mique et multidimensionnel du **développement social et affectif**.
5. airage de quelques repères du **développement psychoaffectif** de l'enfant, nous établissons
6. n l'histoire infantile de son **développement psychoaffectif** et de ses relations familiale
7. istic, spatial cognition and **affective development** - illustre this complex pro
8. ackground. Research indicates that **affective aspects of development provide a basis for autonomous learni**

results: 1 - 8

Concept

 **Développement affectif**

Check/Uncheck all

terms



Développement affectif  
 Développement psychoaffectif

terms



Affective development  
 Affective Development

terms



Desarrollo afectivo

related

Trouble du développement ; Motricité ;

broader

Psychologie ;

narrower

Affect affectivité ; Aptitude sociale ; Sentiment ; Vulnérabilité ;



# Achievements and perspectives

- FastKwic freely available for the community  
<http://www.cnrtl.fr/outils/>.
- Possible improvements :
  - ▶ integration of new languages
  - ▶ linking FastKwic to a terminology extraction tool (ex. ACABIT)

001. and you, my lady, are the living image of him!" ["THANK YOU](#), my lady." "The duke will never be dead  
 002. if you want. I'm not doing anything important." ["THANK YOU](#)," said Karen. "You're very kind. I've go  
 003. sign at the outskirts of the village that said, ["THANK YOU](#)." "What are you grinning at?" Mary asked  
 004. time, sir. That's about half an hour from now." ["THANK YOU](#)." Carl changed the time on his watch. "A  
 005. "Dunno about tea; fuckin' good at makin' a noise." ["THANK YOU](#) for sharing that with us, Gav. I shall r  
 006. , get out. I'm a busy man." "Goodbye," said Kee. ["THANK YOU](#) for talking to me. But don't forget, Mr  
 007. and a strong wind was blowing. She turned round. ["THANK YOU](#), gentlemen. I will have to talk to them.  
 008. d, "Shall I put this straight into your basket?" ["THANK YOU](#) so much!" She would never be able to exp  
 009. d: "What shall I say? Nobody could foresee this." ["THANK YOU](#) for this wonderful night, an emotional W  
 010. ed you would call on us." Joan hid her surprise. ["THANK YOU](#) for the letter, Your Grace -- 'twas kind  
 011. en me so much happiness. I buy her jewels to say ["THANK YOU](#)." In May 1972 the Duke became ill. Whe

## THANK YOU FOR YOUR ATTENTION

012. Here you are." Harald gave the man his passport. ["THANK YOU](#). And his?" "He has no passport. I am a p  
 013. ince could in the fullness of time marry... ["THANK YOU](#) for telling me, my lady," Joan said. "I  
 014. ing was exquisitely ornamented with tiny pearls. ["THANK YOU](#), my lord," she said, her eyes shining wi  
 015. ive us this day our daily bread", and seldom say ["THANK YOU](#)" for the ways that we see answers to the  
 016. ker laughed again and moved on to the next seat. ["THANK YOU](#), Harald," Carl whispered, when the man  
 w  
 017. ld me that two of those men had come back to say ["THANK YOU](#) and take Deputy Superintendent Dr Lloy  
 d  
 018. little longer. It was very nice to talk to you." ["THANK YOU](#) for talking to me, too. I've learnt a lo