


AFCP

ATALA

Journées d'Études sur la Parole
Traitement Automatique des Langues Naturelles
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues

# PARIS Inalco du 4 au 8 juillet 2016

## Organisé par les laboratoires franciliens

https://jep-taln2016.limsi.fr

## Conférenciers invités:

Christian Chiarcos (Goethe-Universität, Frankfurt.)
Mark Liberman (University of Pennsylvania, Philadelphia)

## Coordinateurs comités d'organisation

Nicolas Audibert et Sophie Rosset (JEP)
Laurence Danlos & Thierry Hamon (TALN)
Damien Nouvel & Ilaine Wang (RECITAL)
Philippe Boula de Mareuil, Sarra El Ayari & Cyril Grouin (Ateliers)

©2016 Association Francophone pour la Communication Parlée (AFCP) et
Association pour le Traitement Automatique des Langues (ATALA)

# Table des matières

*ACGTK : un outil de développement et de test pour les grammaires catégorielles abstraites*
Sylvain Pogodalla . . . . . 1

*Un analyseur de conversations pour la relation client*
Hugues de Mazancourt, Gaëlle Recourcé, Soufian Salim . . . . . 3

*AppFM, une plate-forme de gestion de modules de TAL*
Paul Bui-Quang, Brigitte Grau, Patrick Paroubek . . . . . 6

*CommunicoTool Advance, un prototype d'application d'aide à la communication*
Charlotte Roze . . . . . 9

*E-Quotes : un outil de navigation textuelle guidée par les annotations sémantiques*
Motasem Alrahabi . . . . . 12

*Exploration de collections d'archives multimédia dans le contexte des Humanités Numériques : revisiter TALN'2015 ?*
Géraldine Damnati, Marc Denjean, Delphine Charlet . . . . . 15

*FlexiMac 1.1. - Conjugueur automatique des verbes macédoniens*
Jovan Kostov . . . . . 18

*Héloïse, une plate-forme pour développer des systèmes de TA compatibles Ariane en réseau*
Vincent Berment, Christian Boitet, Guillaume de Malézieux . . . . . 21

*Identification de lieux dans les messageries mobiles*
Clément Doumouro, Adrien Ball, Joseph Dureau, Ramzi Ben Yahya, Sylvain Raybaud . . 24

*Interface Web pour l'annotation morpho-syntaxique de textes*
Thierry Hamon . . . . . 26

*Lecture bilingue augmentée par des alignements multi-niveaux*
François Yvon, Yong Xu, Marianna Apidianaki, Clément Pillias, Pierre Cubaud . . . . . 29

*LNE-Visu : une plateforme d'exploration et de visualisation de données d'évaluation*
Guillaume Bernard, Olivier Galibert, Rémi Régnier, Séverine Demeyer, Juliette Kahn . . . 31

*Un outil multilingue d'extraction de collocations en ligne*
Luka Nerima, Violeta Seretan, Eric Wehrli . . . . . 34

*Radarly : écouter et analyser le web conversationnel en temps réel*
Jade Copet, Christine de Carvalho, Virginie Mouilleron, Benoit Tabutiaux, Hugo Zanghi . 37

*SOFA : Une plateforme d'analyse syntaxique en ligne pour l'ancien français*
Gaël Guibon . . . . . 40

*STAM : traduction des textes non structurés (dialectes du Maghreb)*
Mehdi Embarek, Soumya Embarek . . . . . 43

*Un système automatique de sélection de réponse en domaine ouvert intégrable à un système de dialogue social*
Franck Charras, Guillaume Dubuisson Duplessis, Vincent Letard, Anne-Laure Ligozat, Sophie Rosset . . . . . 46

*Tag Thunder : plateforme de démonstration et d'expérimentation*
Jean-Marc Lecarpentier, Elena Manishina, Maxence Busson, Fabrice Maurel, Stéphane Ferrari . . . . . 49

# FlexiMac 1.1. - Conjugueur automatique des verbes macédoniens

Jovan Kostov[1]
(1) PLIDAM EA 4514 – INALCO, 2, rue de Lille, 75007 Paris, France
jovan.kostov@gmail.com

## RÉSUMÉ

Cette démonstration présente la plateforme FlexiMac 1.1., générateur automatique des verbes macédoniens qui permet de conjuguer un verbe dans la plupart des modes et des temps, sans faire appel à une base de données. Après un bref exposé du fonctionnement de la plateforme, nous allons également évoquer les travaux actuels qui en ont découlé, et ceux qui sont en train d'émerger dans une perspective de traitement automatique du macédonien en tant que langue européenne peu-dotée.

## ABSTRACT

**FlexiMac 1.1. – Automatic Generator of the Verbal Paradigms of the Macedonian language.**
This paper presents the FlexiMac 1.1. platform, an automatic generator of Macedonian verbs that conjugates verbs in most moods and tenses, without any use of a database. After a brief presentation of the platform's architecture, we will also discuss the current and emerging works that resulted from this research and that can be used for a natural processing of the Macedonian language as one of the less-documented European languages.

MOTS-CLÉS : macédonien, verbe, flexion, génération, conjugaison, plateforme, web.
KEYWORDS: Macedonian, verb, inflection, generation, conjugation, platform, web.

# 1 Qu'est-ce que FlexiMac 1.1. ?

FlexiMac 1.1. est une plateforme-web conçue en utilisant le langage PHP. La plateforme-web est ouverte, hébergée à l'adresse http://fleximac.free.fr. Il s'agit d'un système non-supervisé de traitement des chaînes de caractères (les formes lexicales d'un verbe) envoyées à travers des formulaires. La plateforme génère les paradigmes d'un mode ou d'un temps verbal à travers les étapes suivantes :

- l'utilisateur saisit un verbe dans un formulaire ;
- FlexiMac 1.1. analyse le verbe et détermine les opérations (allomorphies) qu'il doit effectuer sur le thème verbal, puis lui accole les terminaisons adéquates ;
- la plateforme affiche les formes du verbe dans un tableau.

## 1.1 Architecture algorithmique et contraintes linguistiques

Les algorithmes de FlexiMac 1.1. sont des automates déterministes qui permettent tout d'abord une reconnaissance du groupe auquel appartient un verbe. Suivant le mode et le temps demandé, ils effectuent une transformation et un réajustement morphologique du thème. La génération des formes à proprement parler est la troisième et dernière étape qui permet de stocker les formes verbales d'un paradigme dans des variables et de les afficher à l'écran. FlexiMac 1.1. n'utilise aucune base de données.

L'utilisateur final n'est pas obligé d'avoir une littératie numérique spécifique et la seule procédure se résume à la saisie d'un verbe dans le formulaire et à l'affichage des résultats dans des tableaux contenant des informations sur le nombre, le genre et la personne. Une illustration-video exhaustive du fonctionnement de FlexiMac 1.1. peut être consultée à l'adresse http://fleximac.free.fr/fra/faq2.php.

### 1.2 Apport critique de la communauté et modélisation interactive de la flexion

FlexiMac 1.1. est une plateforme ouverte et vise principalement les apprenants francophones et les locuteurs natifs du macédonien. Sur les 25000 verbes testés, la plateforme conjugue avec une précision de 91 %. Les autres verbes (9 %) sont des verbes qui ont une conjugaison flottante qui peut être caractérisée comme irrégulière. Pour certains paradigmes verbaux, la forme diffère considérablement de l'usage et c'est pour cela que nous avons entrepris un travail collaboratif au sein du groupe Facebook « Puzder pravam od makedonskiot jazik[1] » qui compte 15000 membres, essentiellement des locuteurs natifs, mais aussi des professionnels de la langue (linguistes, professeurs de macédonien, traducteurs et relecteurs professionnels). Les utilisateurs donnent un feedback régulier pour améliorer le fonctionnement FlexiMac 1.1. et pour pointer les zones d'ombre de la conjugaison macédonienne, ce qui représente un apport considérable à la description de la flexion de cette classe de mots par la communauté-même.

## 2 Perspectives : linguistique descriptive et TAL

La modélisation informatique du verbe macédonien et la programmation de FlexiMac 1.1. nous a permis de nous interroger sur les règles de la flexion verbale qui n'étaient pas utilisées dans les grammaires existantes de référence (Koneski, 1952/2005 ; Minova-Gjurkova, 2006). En effet, seules quelques études (Friedman, 1977 ; Elson, 1989) abordaient cette question. De ce fait, au lieu de recourir à un bricolage, nous avons systématisé les règles de la flexion. Cette recherche nous a permis de repenser entièrement la conjugaison macédonienne et de la rendre accessible à des locuteurs dont les systèmes verbaux natifs (français, espagnol, anglais) fonctionnent d'une manière différente et ne présentent pas une complexité comme celle du macédonien où la présence de l'aspect – comme dans toutes les langues slaves, d'ailleurs – rend l'apprentissage très laborieux.

La plateforme FlexiMac 1.1. fonctionne maintenant depuis 3 ans et demi. De nombreuses améliorations ont été apportées et le graphisme de l'interface épuré et amélioré. Actuellement, nous travaillons à la description des verbes irréguliers et défectifs pour savoir s'il s'agit d'un système « régulier » qui représente une classe flexionnelle à part-entière ou une série d'irrégularités qui ne peuvent pas être insérées dans les cases conventionnelles de la conjugaison macédonienne. L'étape suivante consistera à répertorier toutes les formes fléchies dans une base de données qui permettra l'extraction d'un lexique morphologique pour des travaux ultérieurs sur cette langue comme, par exemple, l'étiquetage morphosyntaxique des corpus.

## Remerciements

Nous tenons à remercier toute personne qui a collaboré de loin ou de près à la réalisation de cette plateforme, et notamment nos informateurs de langue maternelle macédonienne et nos collègues et étudiants de la section de macédonien à l'INALCO. Un remerciement spécial va à l'adresse de toute la communauté Facebook du groupe « Puzder Pravam od makedonskiot jazik » pour sa contribution significative à la modélisation des formes verbales les plus problématiques. Enfin, nous tenons à remercier Stasha Tofoska pour les longues heures consacrées à la réflexion commune sur l'aspect – catégorie qui ne cesse de fasciner tout apprenant d'une langue slave.

[1] https://www.facebook.com/groups/makedonskijazik/

# Références

ELSON M.-J. (1989). *Macedonian verbal morphology*, Bloomington : Slavica Publishers.

FRIEDMAN V.-A. (1977). *The Grammatical Categories of the Macedonian Indicative*, Bloomington : Slavica Publishers.

KONESKI B. (1952/2005). *Gramatika na makedonskiot literaturen jazik*, Skopje : Prosvetno Delo.

KONESKI K. (1999). *Pravopisen re nik na makedonskiot jazik*, Skopje : Prosvetno Delo.

KOSTOV J. (2013). *Le verbe macédonien : pour un traitement automatique de nature linguistique et applications didactiques (réalisation d'un conjugueur)*, thèse de doctorat soutenue le 7 décembre 2013, Paris, INALCO.

MINOVA-GJURKOVA L. (2006). *Grammaire de la langue macédonienne pour étrangers*, Štip : 2-ri avgust.

Puzder pravam od makedonskiot jazik – Groupe Facebook : http://www.facebook.com/groups/makedonskijazik (dernière consultation le 26 mai 2016).

Tutoriel vidéo de fonctionnement de FlexiMac 1.1. : http://fleximac.free.fr/fra/faq2.php (dernière consultation le 26 mai 2016).