

OenoLex Bourgogne : un corpus oral polyvalent

Laurent Gautier, EA 4182, Univ. Bourgogne
Franche-Comté & USR 3516

Hédi Maazaoui, USR 3516

Patrick Leroyer, Centre for lexicography,
Aarhus University

Structure de la présentation

1. Objectifs
2. Éléments de contextualisation
3. Constitution du corpus
4. Traitement et niveaux d'annotation
5. Exploitation lexicographique en cours
6. Perspectives

1. Objectifs

- Présentation d'un corpus inédit d'interactions de dégustation : OenoLex Bourgogne
- Présentation et discussion des choix méthodologiques opérés dans une perspective dynamique
- Présentation et discussion des potentialités d'exploitation du corpus oral en lexicographie spécialisée et au-delà

2. Éléments de contextualisation

OenoLex Bourgogne : un projet interdisciplinaire



- Une demande de la filière professionnelle
- Objectifs : construction d'un outil de description, représentation et communication des connaissances œnologiques sur les vins de Bourgogne à *destination* des consommateurs – situations didactiques dégustation et promotion
- Impératifs du commanditaire :
 - Sortir du discours expert
 - Partir de situations de dégustation *authentiques*
 - Créer un outil innovant par rapport aux BD ou glossaires terminologiques traditionnels

- Impératifs pour les chercheurs :
 - Accéder à un nouveau type de données en LSP (du vin) : données orales (vs. notes de dégustation) spontanées (vs. dégustation de concours) => les « mettre en production »
 - Concevoir des outils permettant d’appréhender les « dysfonctionnements » terminologiques des consommateurs
 - Représenter l’information générée de la façon la plus précise *et* conviviale pour l’utilisateur final

3. Constitution du corpus

- Enregistrements « authentiques » en situation de dégustation asymétrique : animateur-expert vs. groupe consommateurs/ visiteurs

Situation	Nombre d'enregistrements	Durée
Didactique	4	4h33min20s
Marketing direct	3	2h9min5s
Marketing promotionnel et PR	5	1h33min19s
	12	8h15min44s

- Niveau technique :
 - plateforme ADN de MSH Dijon
 - traitement conforme aux préconisations standard pour corpus oraux (SIAF, CINES, TGIR HUMA-NUM, consortium IRCOM)
- Corpus ouvert (autres appellations, puis autres régions) et évolutif (entre autres grâce à la souplesse de Sonal) => OenoLex devenu une sorte de « corpus tout-en-un »

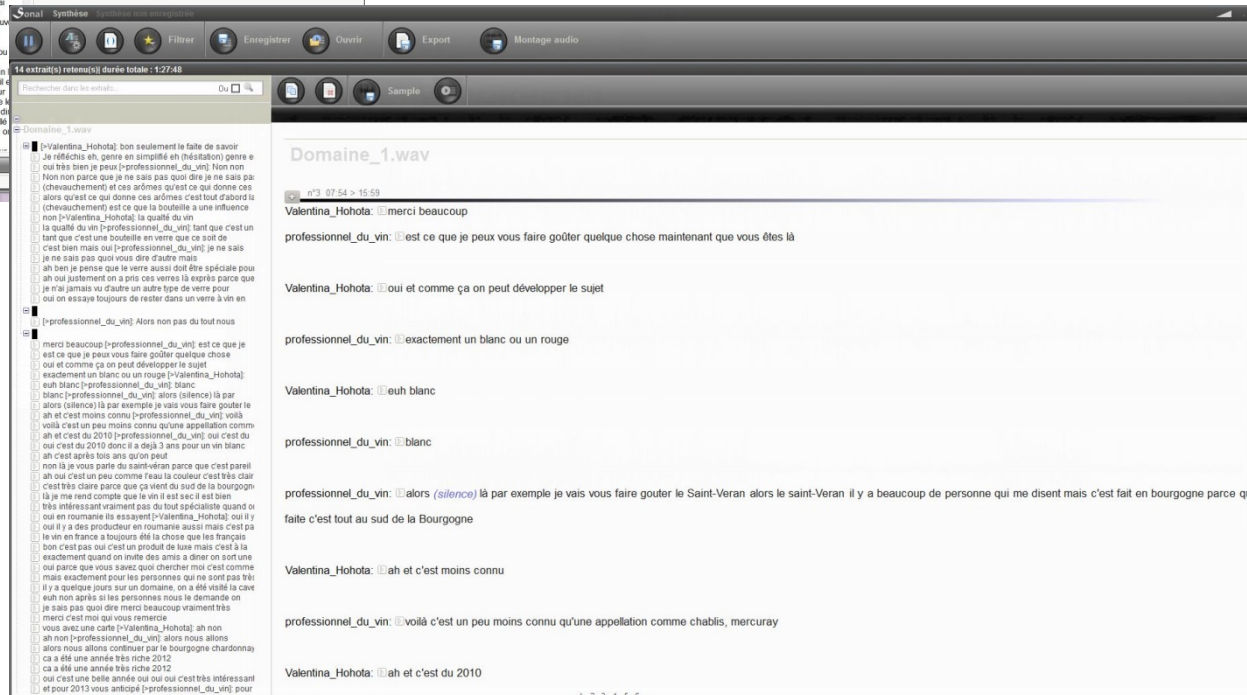
4. Traitement et niveaux d'annotation

- Transcription orthographique + productions orales non-verbales + éléments situationnels

[domaine_chevrot 7'01" à 7'25"]

[professionnel] : alors nous allons continuer la la dégustation pareil on reste sur un bourgogne chardonnay cette fois-ci mais en haute côte de beaune en millésime 2012 également [bruit de verre] donc là on va entrer dans un peu plus de complexité on a des des origines de sols qui vont être un peu plus argilo-calcaires et argilo-calcaires grasseux donc on va se rendre compte qu'il y a un peu plus de rondeur [silence] tout de suite c'est un peu plus rond

• Alignement sous Sonal avec le flux audio



- Exploitation des fonctions d'indexation de l'outil pour balisage du contenu => double sortie qualitative et quantitative
- Choix d'une description maximaliste permettant d'envisager des exploitations par delà l'objectif premier => 3 niveaux d'annotation

- 1^{er} niveau macrolinguistique : actes de langage réalisés
 - Conditions de « gestion » de l'interaction entre expert et consommateur
 - Mise en évidence la dimension hédonique (exclamations, verbalisation d'émotions,...)

- 2^{ème} niveau : description du « contenu »
 - Choix de descripteurs élaboré au fil de la transcription => thématiques traversant le corpus
 - Descripteurs hiérarchisés :
 - Histoire du domaine
 - Présentation du vin
 - Cépage
 - Millésime
 - Parcelle
 - Evaluation sensorielle
 - Œil
 - Nez
 - Bouche

P **[Définition]** alors sakura ça veut dire fleur de fleur de cerisier

C oui

P **[Définition]** c'est un

C **[Question]** ah c'est c'est un nom japonais disons

P japonais

C **[Question]** ah oui et pourquoi ?

P [chevauchement] plus sakura [rires] sakura **[Anecdote] [Famille]** parce qu'en fait la propre la la la femme de Pablo Chevrot

C oui

P est japonaise

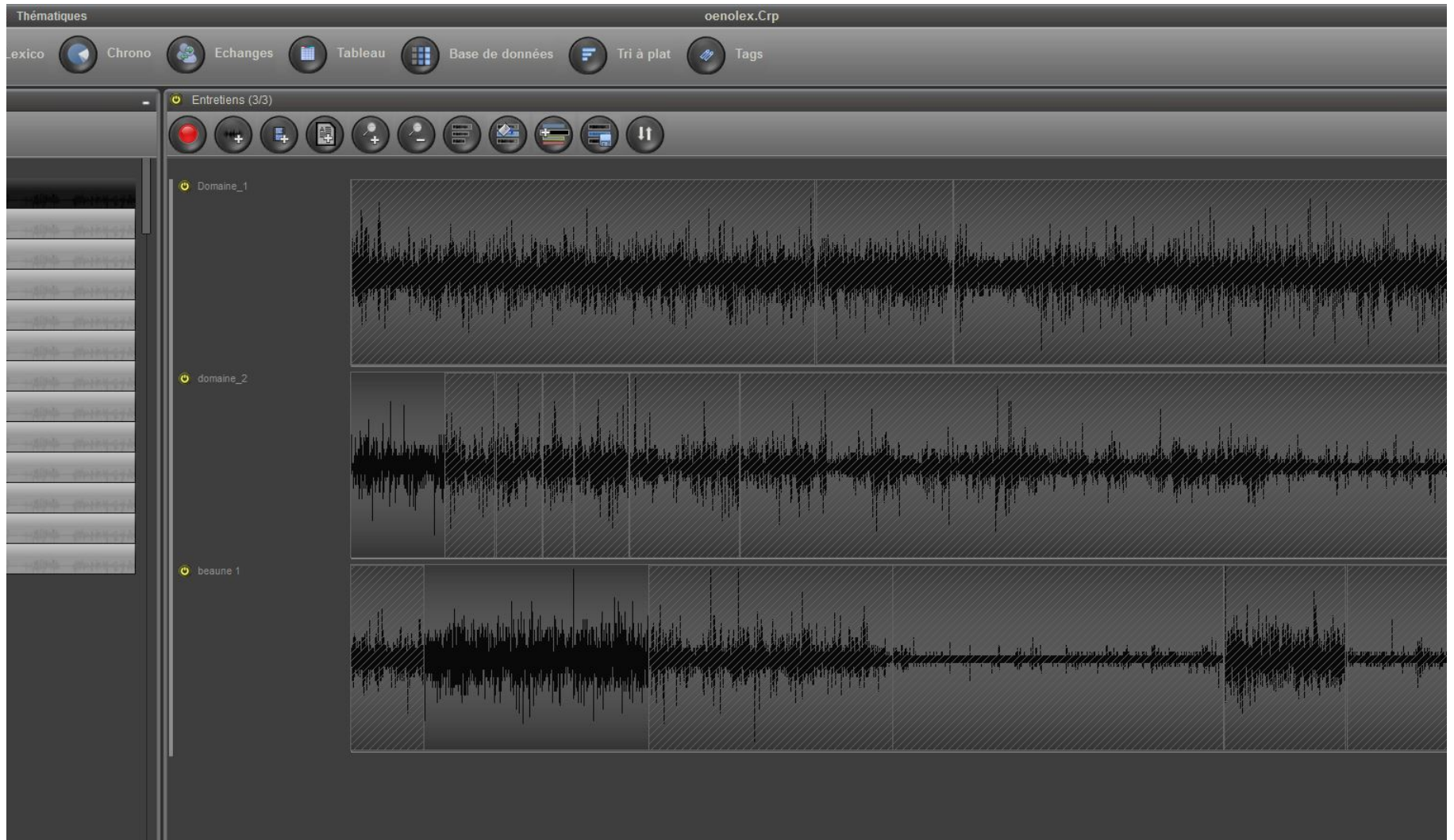
C ah voilà

P **[Présentation]** et donc c'est un rosé qui est principalement dédié au comment euh à la clientèle japonaise asiatique

C [chevauchement] ah

P **[Accord]** pour l'accord mets et vins [Sensorialité] parce que les les rosés de Bourgogne sont un peu plus poivrés en fin de bouche

- Double intérêt :
 - Esquisse du « moule interactionnel » de la dégustation (macrostructure)
 - Premier balisage pour l'exploitation lexicographique
 - permettant d'attribuer du lexique (et des structures) aux différentes phases de la dégustation
 - permettant aussi de repérer, en fonction des tours de paroles, la circulation et l'appropriation ou non de la terminologie de l'expert (cf. « dysfonctionnements »)



- 3^{ème} niveau : marqueurs pour l'exploitation lexicographique
 - But : isoler les unités pertinentes pour l'outil commandé
 - Balisage : terme, exemple, contexte, collocation, etc.

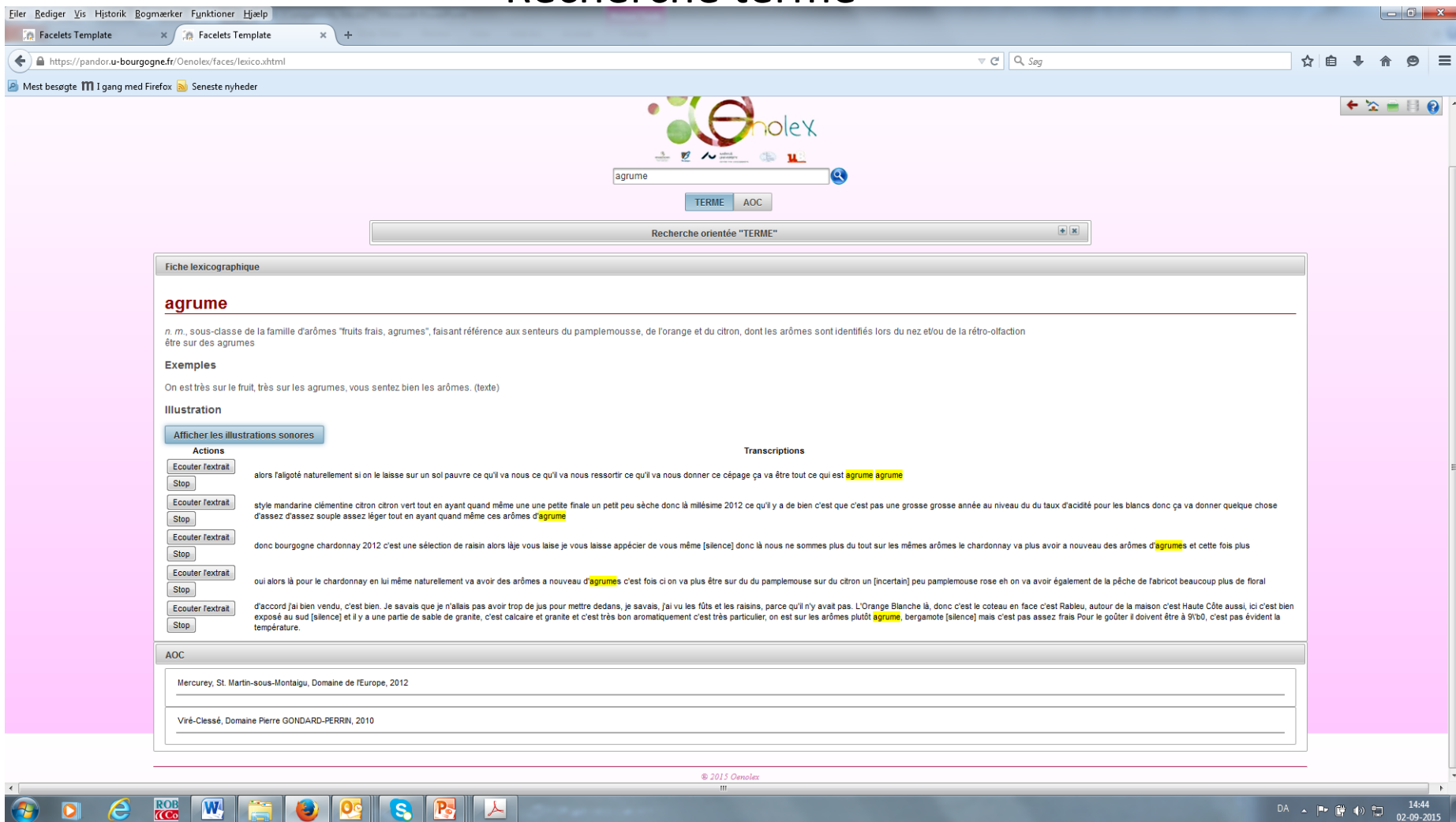
5. Exploitation lexicographique en cours

Extraction automatique des données lexicographiques:

- Extraction lemmatique et extraction des collocations
- Extraction des exemples lexicographiques (= séquences d'enregistrement) au moyen des contextes riches en connaissances:
 - Balisage thématique pour les macro-contextes interactionnels (= séquences longues)
 - Balisage discursif pour les micro-contextes interactionnels (= séquences courtes)

Interface utilisateur double accès: consultation terme ou AOC

Recherche terme



Facelets Template x Facelets Template x

https://pandor.u-bourgogne.fr/Oenolex/faces/lexico.xhtml

Mest besøgte 11 I gang med Firefox Seneste nyheder

Oenolex

agrume

TERME AOC

Recherche orientée "TERME"

Fiche lexicographique

agrume

n. m., sous-classe de la famille d'arômes "fruits frais, agrumes", faisant référence aux senteurs du pamplemousse, de l'orange et du citron, dont les arômes sont identifiés lors du nez et/ou de la rétro-olfaction être sur des agrumes

Exemples

On est très sur le fruit, très sur les agrumes, vous sentez bien les arômes. (texte)

Illustration

Afficher les illustrations sonores

Actions Transcriptions

Ecouter l'extrait Stop alors taligoté naturellement si on le laisse sur un sol pauvre ce qu'il va nous ce qu'il va nous ressortir ce qu'il va nous donner ce cépage ça va être tout ce qui est **agrume agrume**

Ecouter l'extrait Stop style mandarine clémentine citron citron vert tout en ayant quand même une petite finale un petit peu sèche donc là millésime 2012 ce qu'il y a de bien c'est que c'est pas une grosse grosse année au niveau du taux d'acidité pour les blancs donc ça va donner quelque chose d'assez d'assez souple assez léger tout en ayant quand même ces arômes d'**agrume**

Ecouter l'extrait Stop donc bourgogne chardonnay 2012 c'est une sélection de raisin alors làje vous laisse je vous laisse apprécier de vous même [silence] donc là nous ne sommes plus du tout sur les mêmes arômes le chardonnay va plus avoir a nouveau des arômes d'**agrumes** et cette fois plus

Ecouter l'extrait Stop oui alors là pour le chardonnay en lui même naturellement va avoir des arômes a nouveau d'**agrumes** c'est fois ci on va plus être sur du du pamplemousse sur du citron un [incertain] peu pamplemousse rose eh on va avoir également de la pêche deabricot beaucoup plus de floral

Ecouter l'extrait Stop d'accord j'ai bien vendu. c'est bien. Je savais que je n'allais pas avoir trop de jus pour mettre dedans, je savais, j'ai vu les fûts et les raisins, parce qu'il n'y avait pas. L'Orange Blanche là, donc c'est le coteau en face c'est Rableu, autour de la maison c'est Haute Côte aussi, ici c'est bien exposé au sud [silence] et il y a une partie de sable de granite, c'est calcaire et granite et c'est très bon aromatiquement c'est très particulier, on est sur les arômes plutôt **agrume**, bergamote [silence] mais c'est pas assez frais Pour le goûter il doivent être à 9/10, c'est pas évident la température.

AOC

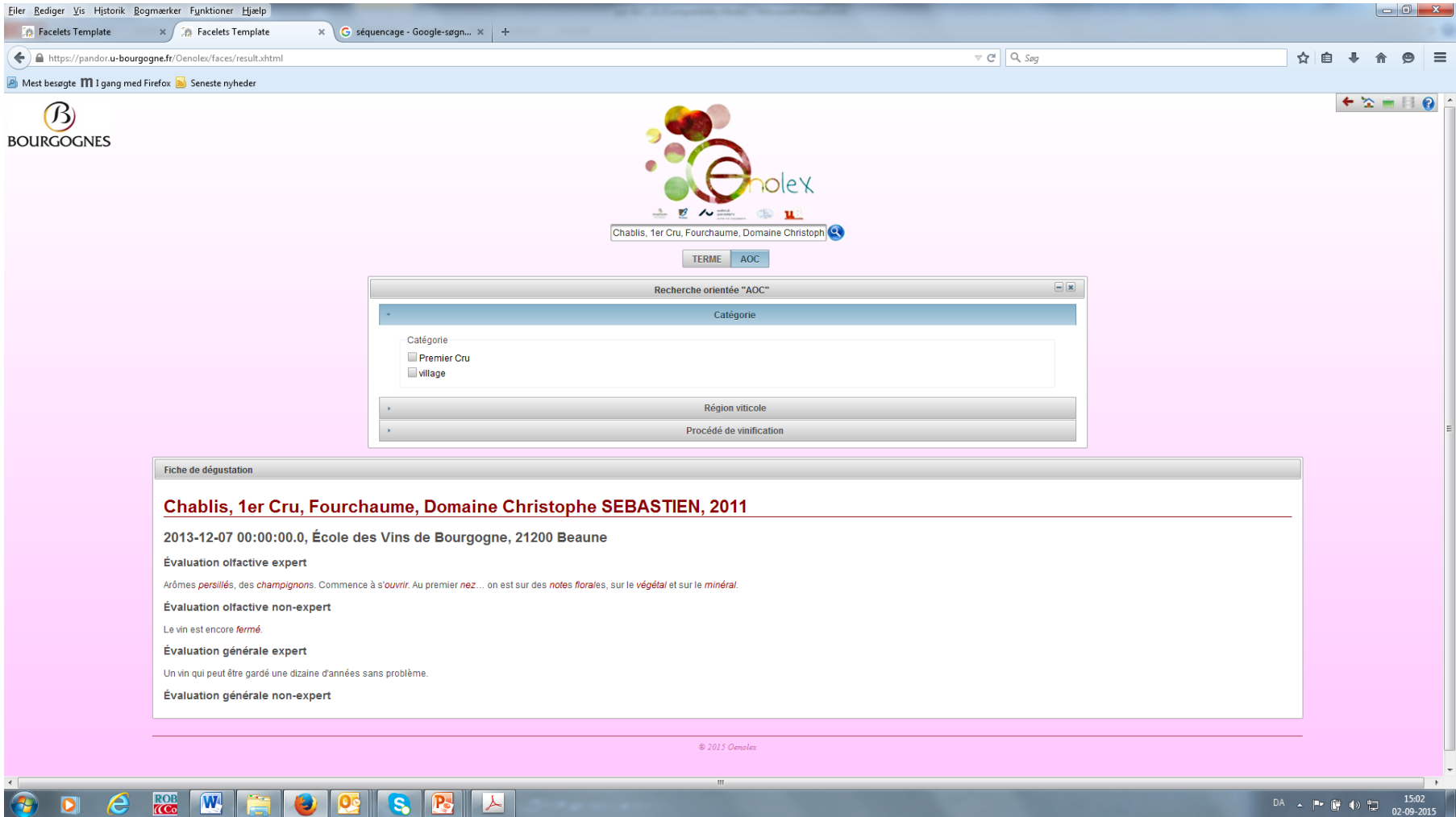
Mercury, St. Martin-sous-Montaigu, Domaine de l'Europe, 2012

Viré-Clessé, Domaine Pierre GONDARD-PERRIN, 2010

© 2015 Oenolex

14:44 02-09-2015

Recherche AOC



Facelets Template x Facelets Template x séquencecage - Google-sagn... x +

https://pandor.u-bourgogne.fr/Oenolex/faces/result.xhtml

Mest besøgte 11 gang med Firefox Seneste nyheder

BOURGOGNES

Oenolex

Chablis, 1er Cru, Fourchaume, Domaine Christoph

TERME AOC

Recherche orientée "AOC"

Catégorie

Catégorie

Premier Cru

village

Région viticole

Procédé de vinification

Fiche de dégustation

Chablis, 1er Cru, Fourchaume, Domaine Christophe SEBASTIEN, 2011

2013-12-07 00:00:00.0, École des Vins de Bourgogne, 21200 Beaune

Évaluation olfactive expert

Arômes *persillés*, des *champignons*. Commence à *s'ouvrir*. Au premier nez... on est sur des *notes florales*, sur le *végétal* et sur le *minéral*.

Évaluation olfactive non-expert

Le vin est encore *fermé*.

Évaluation générale expert

Un vin qui peut être gardé une dizaine d'années sans problème.

Évaluation générale non-expert

© 2015 Oenolex

15:02 02-09-2015

Système de management des données

ENVIRONNEMENT DÉVELOPPEMENT

- Langage: Java v8, XML and XSL(T)
- IDE: NetBeans v8.0.2
- Frameworks : Java Server Faces v2.2, Prime Faces v5.x

SERVEUR

- Linux Debian 7 (Wheezy) 64 bits
- Serveur Apache Tomcat 8.x
- Base de données RDBMS : MySQL 5.x

Extraction des exemples lexicographiques par
séquençage et marquage discursif – verbes process, locatifs,
impératifs, négatifs, ...

On va ***continuer*** la dégustation par un Bourgogne Aligoté 2012....;

Au ***premier nez, on est sur*** des fruits plus mûrs, on n' est pas sur le
côté citronné, agrume, mais sur le côté prune.

le vin a vieilli, mais il a vieilli combien ? C' est ça la question, ***chauffez-
le bien dans vos mains.***

On n' est ***pas*** sur le côté citron, agrume, ***mais*** sur le côté prune.

Exemples lexicographiques - découpage séquentiel = verbes de guidage et débrayage/embrayage discursif

[*On va continuer la dégustation par un Bourgogne Aligoté 2012....*]

[*Pour en revenir un peu sur l' historique, la philosophie du domaine plutôt...*]

[*On continue avec ce qu' on appelle Cuvée Prestige*]

[*pareil on reste sur un bourgogne chardonnay cette fois-ci*]

Exemples lexicographiques - découpage séquentiel: ancrage/balisage sensoriel

[Au **premier nez, on est sur** des fruits plus mûrs, on n' est pas sur le côté citronné, agrume, mais sur le côté prune.]

[Au **deuxième nez, on évolue sur** des arômes un peu miellés, on déduit même une sucrosité – mais c' est en bouche qu' on identifie ça.]

Exemples lexicographiques - didactique instructionnelle

[Nuances de framboise, la couleur est différente, le vin a vieilli, mais il a vieilli combien ? C' est ça la question, ***chauffez-le bien dans vos mains.***]

Exemples lexicographiques - didactique ludique: jeux, questions, devinettes

[Le vert indique la jeunesse ou plutôt l'origine ?]

Exemples lexicographiques - négation = didactique de la recatégorisation de la connaissance face à la non-connaissance

[Au premier nez, on est sur des fruits plus mûrs, **on n' est pas** sur le côté citron, agrume, mais sur le côté prune. Au deuxième nez, on évolue sur des arômes un peu miellés, on déduit même une sucrosité. Quand vous l' aérez, il est assez frais, **c' est pas compoté**, cuit, il y a même aussi un côté floral, **il n' y a pas que le fruit**. On est sur des fruits plus mûrs, **peut-être la pêche**, la prune.]

6. Perspectives

- un nouveau type de corpus en LSP partant de la sphère discursive d'une filière
- un nouveau type de dictionnaire multimodal spécialisé pour professionnels avec implication directe des utilisateurs
- une ressource utilisable pour d'autres applications scientifiques et pratiques, *cf.* projet ADWINE – système de recommandation intelligent avec LE2i

Merci pour votre attention !

Laurent Gautier

EA 4182, Univ. Bourgogne Franche-Comté & USR 3516

laurent.gautier@u-bourgogne.fr

Hédi Maazaoui, USR 3516

hedi.maazaoui@u-bourgogne.fr

Patrick Leroyer, Centre for lexicography, Aarhus University

pl@asb.dk