

Construction et exploitation de ressources numériques sur la vigne et le vin

16 décembre 2016 – Journée d'étude de lancement de projet

«Le vin de l'autre»

Université de Haute-Alsace, Mulhouse

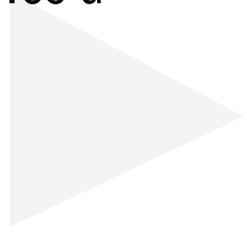
Laurent Gautier, Centre Interlangues Texte Image Langage (UBFC, EA 4182)

Hedi Maazaoui, Maison des Sciences de l'Homme de Dijon (USR CNRS-uB 3516)





1. Objectifs

- Présenter l'**écosystème de recherche et de la valorisation** sur Vigne et Vin sur le grand campus dijonnais avec un focus sur les SHS
 - Présenter un tour d'horizon des **ressources numériques** constituées et de leurs enjeux pour la recherche en termes **méthodologiques**
 - Discuter quelques **problématiques de recherche** possibles à l'intersection entre **sciences du langage** et **info-com**
- 
- 

2. Eléments de contexte (1)

- Un écosystème local et régional au service et / ou en appui à la recherche sur Vigne et Vin incluant les SHS :

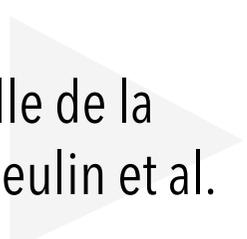


Donnons goût à l'innovation !





2. Éléments de contexte (2)

- Plusieurs projets achevés ou en cours autour des **discours** professionnels de la filière viti-vinicole reposant sur des **corpus numériques** :
 - Projet européen **VinoLingua** à visée didactique (D, E, F, I) (Lavric 2013, Gautier 2014)
 - Projet régional **OenoLex Bourgogne** à visée lexicographique et cognitive – corpus oral (F, extensions plurilingues en cours de négociation) (Gautier/Leroyer 2016)
 - Projet Interreg IV F-CH : construction sémantique et sensorielle de la **minéralité** (F) – triple corpus : écrit, oral, expérimental (Deneulin et al. 2014)
- 
- 

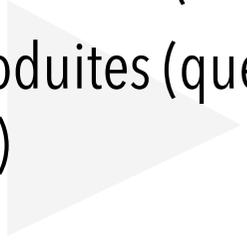


2. Éléments de contexte (3)

- Projet BSN autour du **Bulletin de l'Organisation Internationale de la Vigne et du Vin** (Gautier 2015) (D, E, F, GB, I)
 - Projet **Bibliothèque numérique** de l'OIV (F)
 - Projet **Crémant de Bourgogne** en lien avec UPECB (F) – triple corpus (Mancebo-Humbert/Gautier 2015)
 - Projet **Mémoires du vin** avec la communauté de communes de Gevrey-Chambertin (F)
 - Projet **Adwine** de recommandation consos (lauréat CVT-Athena) (F)
- => Synthèse dans une offre globale de service **Words4Wine**

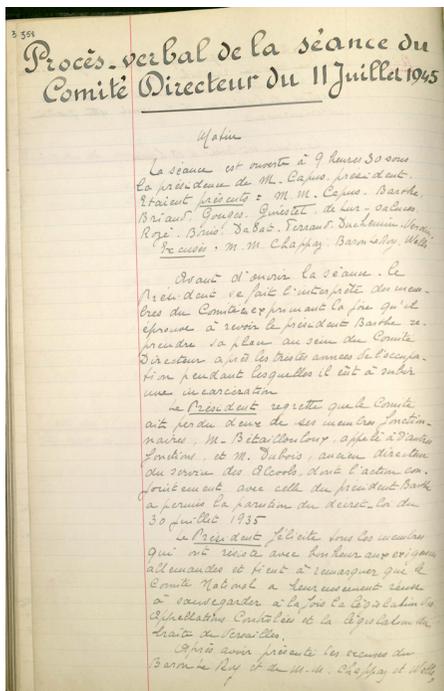


3. Typologie des ressources

- Différents modes de typologisation possibles
 - Par **type de données** :
 - Textes
 - Images fixes + données géospatialisées
 - Données audio(-visuelles)
 - Corpus de « paroles » (*cf. infra*)
 - Par **origine** :
 - Données existantes (bibliothèques, archives, revues)
 - Données produites (questionnaires, entretiens, données de dégustation)
- 
- 

4. Aperçu des ressources disponibles (1)

- Vers une **bibliothèque numérique** de la Vigne et du Vin :
 - Le fonds **INAO** :



Documents de travail et compte-rendu des séances des différents Comités Nationaux depuis leur création (1935-2006)

Bulletins du CNAO puis de l'INAO (1937-1971)

21 200 pages pour les bulletins (plein texte)

5 000 pages pour les CR

https://pandor.u-bourgogne.fr/ead.html?id=FRMSH021_00005



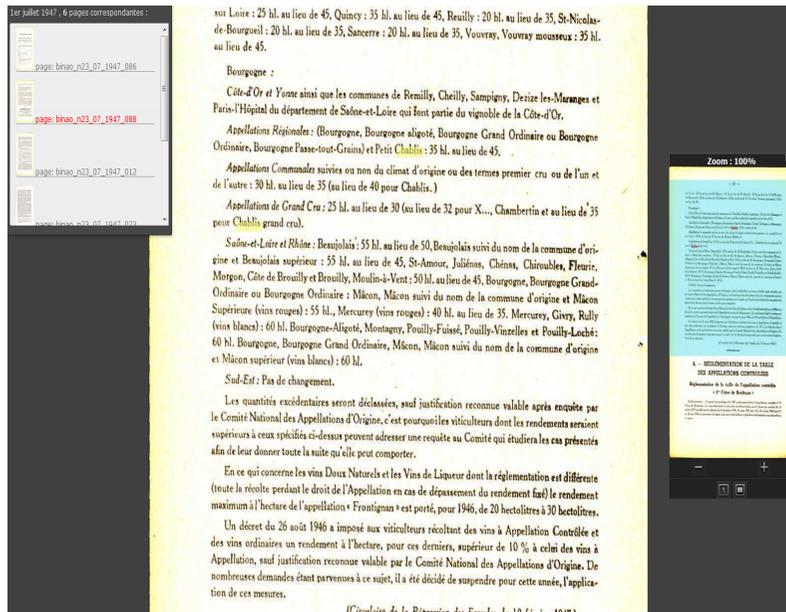
4. Aperçu des ressources disponibles (2)

- La bibliothèque du fonds **OIV** (en cours d'extension) :
 - 7 atlas géographiques
 - 145 ouvrages ayant trait à la thématique de la Vigne et du Vin soit 28 500 pages depuis le XIXème siècle
 - Classement thématique (O. Jacquet) : Découverte des vignobles, Enquêtes sur les vignobles, Evènement professionnel, Organisations professionnelles, Production et Economie, Raisin, Réglementation, Santé / Hygiène, Viticulture/Viniculture
- 
- 

4. Aperçu des ressources disponibles (3)

– Le Bulletin de l'OIV depuis 1928 :

- Volume : 98 000 pages / 5.4 To
- Numérisation => fichiers images + pdf



production de fichiers normalisés XML-ALTO et XML-METS permettant une recherche en plein texte depuis portail PANDOR

création de fichiers TXT pour travail sur corpus

4. Aperçu des ressources disponibles (4)

- Des corpus audio(-visuels) de **discours professionnels**
 - Les enregistrements de **dégustation** (OenoLex, Adwine) :
 - Authentiques
 - Différentes situations de dégustation
 - Flux audio + transcription alignée + annotations sous [®]SONAL

Vue d'ensemble du corpus

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

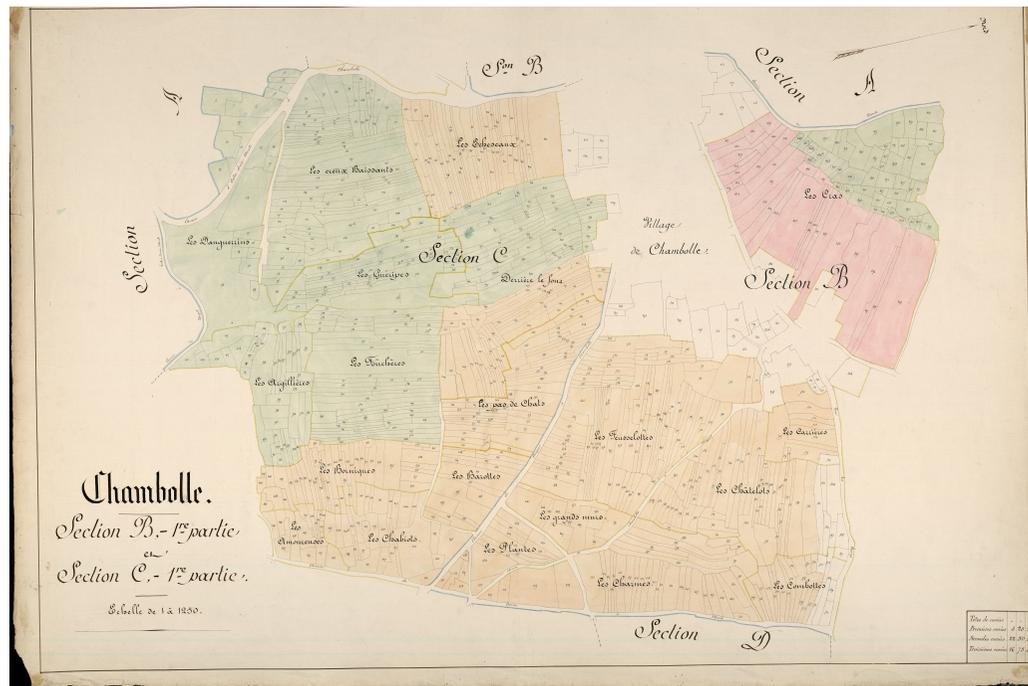
4. Aperçu des ressources disponibles (5)

- Des corpus audio(-visuels) de discours professionnels
 - Les « **mémoires du vin** » de Gevrey Chambertin (12 heures) :
 - Même mode de traitement
 - Visée documentaire et valorisation socio-économique du territoire via une application

[414,8] [>Témoin]: alors quand mon père a vu ça, il dit on peut pas rester là c'était la guerre et puis il savait pas si si il repartirait, si il repartirait pas mais là nan il a dit on restera en France. Alors on a acheté une maison en France mes parents ont achetés une maison en France y z'ont y z'ont voulu s'installer là du coup...

4. Aperçu des ressources disponibles (6)

- Des images fixes (**cartographie**) :
 - 87 cartes, découpage parcellaires des villages de la côte de Beaune et Côte de Nuits

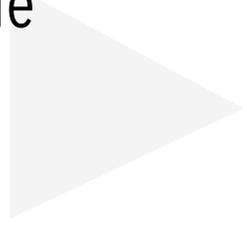


4. Aperçu des ressources disponibles (7)

- Des **corpus de « paroles »** produites en appui à des thèses :
 - Ludivine DOMONT-LEVEQUE : Minéral/minéralité : étude diachronique de la construction d'un descripteur sensoriel dans les textes réglementaires, prescriptifs et descriptifs de la filière vitivinicole (français, allemand, anglais)
 - Corpus trilingue
 - Guides + revues + blogs + forums
 - Mariele MANCEBO-HUMBERT : La construction discursive et terminologique dans la filière viti-vinicole : mieux communiquer auprès des consommateurs et des experts sur le crémant de Bourgogne.
 - Guides + revues + comm. Interne
 - Questionnaires sans stimulus
 - Paroles produites en dégustation



5. Traitement et pistes d'exploitation (1)

- Pas d'exploitation possible sans **traitement informatique, documentaire et sans annotations**
 - Traitement de l'écrit : océrisation + relecture
 - Traitement de l'oral : amélioration des qualités d'enregistrement + transcription + alignement
 - Traitement documentaire : plan de classement + indexation
 - Traitement du contenu : annotation à différentes degrés de granularité
- 
- 

5. Traitement et pistes d'exploitation (2)

- Exemple de traitement oral (mémoires du vin) :

Thématiques

Filter: Et Seulement

Pas de thématique

Sans Thématique

Echanges - Discussions

Parcours

Anecdotes/chansons/proverbes

Anecdotes sur les animaux

Anecdote historique (guerre, résistance etc.)

Anecdote sur la nature

Anecdotes sur la vigne

Anecdotes pittoresques

Sonal 2.0.77 | 09/12/2016 | Gevrey.Crp

Date de copie: 11:36:54

Partiels thématiques: 10:56:07 (94,1%)

Nombre d'extraits: 303

Thématiques (13/15)

- Pas de thématique
- Sans Thématique
- Echanges - Discussions
- Parcours
- Pas de thème
- Anecdotes/chansons/proverbes
- Anecdotes sur les animaux
- Anecdotes historiques (guerre, résistance etc.)
- Anecdote sur la nature
- Anecdotes sur la vigne
- Anecdotes pittoresques
- Chansons - Poèmes
- Proverbes - Dictons - Maximes - Expressions
- Indications géographiques
- Aires viticoles

Entretiens (35/35)

- Mme Marie-Jeanne GALLET
- M. et Mme Eliane et Daniel LORIOT
- M. Camille BAILLY
- M. et M. Françoise (Françoise) & Jules BERNARD
- M. Robert GELLENECOURT
- M. Mme Hervé et Joëlle HEUDELOT
- Mme Bernadette VIN

Tags

Adolescence, Appas, Agriculture, Aire viticole de Gevrey Chambertin, Albert Camus ("Le Premier Homme"), Akékeze (champignon), Amanite panthère, Amanite tue-mouche, Ambiance vendanges, Ambié, Amour en cage (champignon), Anourettes, Anecdote menstruation-cuivrie, Anecdote sur le tee-chaï, Anecdote sur les gens du voyage, Anecdote sur les miliciens espagnols vendangeurs, Anecdote sur l'ouillage, Animaux à fourrure, Animaux nuisibles, Appellation "Bonne-mares", Appellation "Chambertin", Appellation "Chambolle", Appellation "Carmes-Chambertin", Appellation "Clos de Bézu", Appellation "Comte de Vogüé", Appellation "Cuvée Charlemagne", Appellation "Domaine Roumier", Appellation "Les Charmes", Appellation "Les Demoiselles", Appellation "Mazoyères-Chambertin", Appellation "Musigny", Appellation "Voise-Romande", Appellation "Grands Echezeaux", Attachement au territoire, Autonomie alimentaire, Astun, Auxois, Auxonne, Bareszai, Battage, Battoir, Beaune, Bécasse (oiseau), Belette, Béry, Bienfaits du travail de la vigne, Biographie du professionnel, Biographie professionnelle, Biscuitte (champignon), Biscuitte (plante), Bizeau, Boisse (Ba) (source - Gevrey-Chambertin), Bordeaux (vins), Bordeaux (région viticole), Bouille bordelaise, Bourgogne (vins), Braconnage, Brechon, Calottes, Cais (oiseau), Cavaud (oiseau), Carrière de vin Impression, Cave coopérative, Carre, Cépéges, Corvilles, Cuvillotte de carreaux, Chablis, Chalon-sur-Saône, Chambovet, Chambole-Musigny, Champignons, Chanson, Chanson "Joyeux Enfants de la Bourgogne", Chanson bourguignonne, Chanson céri de printemps, Chanterelles (champignon), Chants en Bourgogne, Chasse, Chasse (premier contact), Chasse à la bécasse, Chasse à la calie, Chasse au binaire, Chasse au lièvre, Chasse au sanglier, Chasse dans les vignes, Château de Gevrey-Chambertin, Château-Chalon, Châtillonais, Chéridoine, Chesôve, Chorale, Cisaille, Cierres d'eau, Claude Nolot, Claude Nolot (commandant de Napoléon Ier), Clémency, Clos de Vougeot, Clos Saint-Jacques (Gevrey-Chambertin), Clos Vougeot, Coteau-voisinage d'arrettes, Coche (cervin), Colason, Collet (braconnage), Combe à la Serpent, Combe de Lavaux (Gevrey-Chambertin), Combe Lavaux, Combe Lavaux (Four - hameau de Faux), Commerce des sarmets de vigne, Commerce du vin,

Diagramme circulaire :

Catégorie	Nombre	Pourcentage
Parcours	5453	7,3%
Echanges - Discussions	24025	16,5%
Pas de thématique	10626	8,9%
Anecdotes diverses	14219	13,6%
Anecdotes pittoresques	11402	9,9%
Anecdotes sur la vigne	23658	20,9%
Anecdotes sur la nature	11808	10,4%
Anecdotes historiques (guerre, résistance etc.)	2941	4,0%
Anecdotes sur les animaux	3334	4,5%
Pas transcrite	8303	8,4%
Aires viticoles	8337	8,5%
Proverbes - Dictons - Maximes - Expressions	1118	1,5%
Chansons - Poèmes	1212	1,0%



5. Traitement et pistes d'exploitation (3)

- Approches méthodologiques en cours
 - **Sémantique** lexicale et cognitive : cartographie (interculturelle) des champs de la sensorialité, représentations mentales
 - Traditions **discursives** : émergence et développement des genres propres à la filière selon les langues-cultures
 - Fouille de textes (*text mining*) et **analyse de sentiments** appliquée au vin (*sentiment analysis*)
 - Approches **multimodales** : présence du vin sur les réseaux sociaux
- 
- 



6. Bilan et perspectives

- Ere des données (y compris en SHS)
- Interactions bienvenues avec les différentes disciplines
- Savoir passer de la ressource statique au corpus « intelligent »
- Vraie dynamique offerte par les Humanités Numériques en Europe pour des interconnexions plurilingues et interculturelles (Darjah, Europeana,...)



Merci de votre attention !

Contacts :

Laurent Gautier (laurent.gautier@u-bourgogne.fr)

Hedi Maazaoui (hedi.maazaoui@u-bourgogne.fr)