

# Tutoriel Métadonnées

---

Chaque fichier contient des métadonnées renseignant différentes informations.

Les métadonnées sont renseignées selon le schéma [IMDI](#) sous le logiciel [Arbil](#) .

## Session

Cette composante donne des informations générales sur chaque échantillon.

Pour chaque session sont précisées les informations suivantes:

### Name

Le nom abrégé que l'échantillon a dans le corpus.

Les noms des échantillons Rhapsodie sont codés sur 6 champs, p. ex: [Rhap-M0001, Avanzi] ; [Rhap-D2001, Mertens].

Le premier renseigne sur la nature monologale ('M') ou dialogale ('D') de l'échantillon. Le second indique le domaine privé ('0' ou '1') ou public ('2'). Les 3 suivants, le numéro d'ordre de l'échantillon dans la classe définie par les deux premiers champs. Le dernier champ renseigne la source (interne ou externe) de l'échantillon.

Valeurs :

Rhap-D0001,CFPP2000; Rhap-D0002,CFPP2000; Rhap-D0003,PFC; Rhap-D0004,CFPP2000; Rhap-D0005,PFC; Rhap-D0006,CFPP2000; Rhap-D0007,Avanzi; Rhap-D0008,Avanzi; Rhap-D0009,PFC; Rhap-D0017,Avanzi; Rhap-D0020,Avanzi; Rhap-D1001,ESLO; Rhap-D1002,RHAPSODIE; Rhap-D1003,RHAPSODIE; Rhap-D2001,Mertens; Rhap-D2002,RHAPSODIE; Rhap-D2003,RHAPSODIE; Rhap-D2004,Lacheret; Rhap-D2005,Lacheret; Rhap-D2006,RHAPSODIE; Rhap-D2007,RHAPSODIE; Rhap-D2008,RHAPSODIE; Rhap-D2009,Mertens; Rhap-D2010,RHAPSODIE; Rhap-D2011,RHAPSODIE; Rhap-D2012,RHAPSODIE; Rhap-D2013,RHAPSODIE; Rhap-M0001,Avanzi; Rhap-M0002,Rhapsodie; Rhap-M0003,Avanzi; Rhap-M0004,Avanzi; Rhap-M0005,Avanzi; Rhap-M0006,Avanzi; Rhap-M0007,Avanzi; Rhap-M0008,Avanzi; Rhap-M0009,Avanzi; Rhap-M0010,Avanzi; Rhap-M0012,Avanzi; Rhap-M0013,Avanzi; Rhap-M0014,Avanzi; Rhap-M0015,Avanzi; Rhap-M0016,Avanzi; Rhap-M0017,Avanzi; Rhap-M0018,Rhapsodie; Rhap-M0019,Rhapsodie; Rhap-M0021,Rhapsodie; Rhap-M0022,Rhapsodie; Rhap-M0023,Rhapsodie; Rhap-M1001,RHAPSODIE; Rhap-M1003,RHAPSODIE; Rhap-M2001,C-PROM; Rhap-M2002,RHAPSODIE; Rhap-M2003,RHAPSODIE; Rhap-M2004,RHAPSODIE; Rhap-M2005,RHAPSODIE; Rhap-M2006,RHAPSODIE.

Interrogeable :OUI

## Title

Le nom complet de l'échantillon: il comprend le nom donné dans Rhapsodie, ainsi que le nom que l'échantillon avait dans le corpus source

Interrogeable :NON

## Date

Correspond à la première date de création de l'échantillon. Valeur année sur 4 chiffres.

Interrogeable :NON

## Continent

Le continent où l'échantillon a été enregistré.

Valeurs

- Europe

Interrogeable :NON

## Country

Le pays où l'échantillon a été enregistré ou est originaire.

Valeurs : toujours « France ».

Interrogeable :NON

## Actors

Cette composante regroupe les informations sur l'ensemble des locuteurs de l'échantillon. Pour chaque locuteur sont précisées les informations suivantes :

### Role

Le rôle que le locuteur a dans la session.

Valeurs :

- Researcher: chercheur, locuteur qui élicite la production avec des objectifs de recherche bien définis.
- Interviewer: interviewer, locuteur qui conduit la conversation dans le parlé télévisé ou radiophonique
- Speaker: locuteur, c'est la valeur attribuée à tous les autres locuteurs

Interrogeable : NON

## Name

L'identifiant attribué dans Rhapsodie au locuteur. L'identifiant est constitué de trois informations :

- \$L : désignation d'un locuteur
- M ou F : sexe du locuteur
- Identifiant numérique attribué

Si un locuteur apparaît dans plusieurs échantillon il aura toujours le même identifiant.

Valeurs :

- de \$LF1 à \$LF35 pour les femmes
- de \$LM1 à \$LM52 pour les hommes.

Interrogeable :NON

## Full Name

Le nom complet du locuteur participant dans l'échantillon.

Valeurs:

- Unknown
- Proper Name

Interrogeable :NON

## Age

L'âge du locuteur participant dans l'échantillon.

Valeurs :

- Unknown :inconnu
- Unspecified : non-spécifié
- de 1 à 99 : son âge en chiffre

Interrogeable :OUI

## Sexe

Le sexe du locuteur participant dans l'échantillon.

Valeurs :

- Unspecified: non précisé
- Female: femme
- Male: homme

Interrogeable :OUI

## Language

La langue du locuteur participant à l'échantillon.

Valeurs :

- Français

Interrogeable :NON

## Content

Cette composante donne un ensemble d'informations sur la situation d'énonciation décrite dans l'échantillon.

Pour chaque échantillon sont précisées les informations suivantes :

## Genre

Type de discours de l'échantillon

Valeurs :

- Indiff: non précisé
- Discourse : le contenu se compose d'énoncés produits dans le but de communiquer à son interlocuteur une information ou une intention.
- Ritual/religious text : Le contenu concerne l'accomplissement de rites religieux constitués de types de discours prescrits.

Interrogeable :OUI

## SubGenre

Type de séquence majoritairement représentée dans l'échantillon. Nous nous sommes inspirés, pour définir les valeurs de ce champ, de la théorie des séquences discursives d'Adam (2011), tout en gardant de la classification IMDI la définition du discours oratoire, qui est représentée dans notre corpus.

Valeurs :

- Indiff: non précisé
- Argumentation: construction, par le locuteur, d'une représentation discursive visant à modifier le point de vue de l'interlocuteur sur un objet de discours donné.
- Description : représentation des caractéristiques d'une personne, d'un objet ou d'un événement.
- Procedural : instruction, ou description des étapes impliquées dans l'exécution d'une tâche.

Oratory : discours formel

Interrogeable :OUI

## Interactivity

Caractérise le degré d'interaction entre les locuteurs présents dans l'échantillon.

Valeurs :

- Unspecified: indéfini
- Unknown : non connu
- Interactive : discours caractérisé par un échange entre les locuteurs

- Semi-Interactive : discours ponctué par des interjections répétées de la part de l'interlocuteur
- Non-Interactive : discours caractérisé par de longues séquences que le locuteur produit sans s'attendre à des réponses verbales de la part de son interlocuteur

Interrogeable :OUI

## Planning type

Indique dans quelle mesure le consultant planifie l'événement linguistique.

Valeurs :

- Unspecified: indéfini
- Unknown : non connu
- Spontaneous : discours avec prise de parole libre dont le sujet est librement choisi par les interlocuteurs
- Semi-spontaneous: discours avec prise de parole et sujet dirigé par l'un des locuteurs
- Planned : discours dont le sujet et la structure sont préparés à l'avance

Interrogeable :OUI

## Implication

Indique dans quelle mesure le chercheur est impliqué dans l'événement linguistique.

Valeurs :

- Unspecified: indéfini
- Unknown : non connu
- Elicited: discours provoqué par le chercheur en vue d'obtenir des structures linguistiques précises
- Non-Elicited : discours non élicité

Interrogeable :NON

## Social context

Indique le contexte social dans lequel l'événement prend place.

Valeurs :

- Unspecified: indéfini
- Unknown : non connu
- Private: l'accès à la situation communicative est limité à un nombre restreint d'individus
- Public: l'accès à la situation communicative est ouvert à tous, soit librement soit d'une façon régulée

Interrogeable :OUI

## Event Structure

Indique la structure de l'événement communicationnel.

Valeurs :

- Unspecified: indéfini
- Unknown : non connu
- Monologue: discours produit par un seul locuteur
- Dialogue: discours produit par deux locuteurs
- Conversation: discours produit par deux locuteurs ou plus

Interrogeable :OUI

## Channel

Indique le moyen de communication.

- Valeurs :
- Unspecified: indéfini
- Unknown : non connu
- Face-to-face : la transmission du message est centrée sur l'interaction multi-modale entre le locuteur et ses interlocuteurs
- Broadcasting: le message est transmis à un public large à travers un moyen de communication de masse (radio, télévision, etc.)

Interrogeable : OUI

## Projet

Cette composante regroupe des informations sur le projet Rhapsodie pour son implémentation dans une base de données plus conséquente. Pour cette composante sont précisées les informations suivantes:

### Name

Nom abrégé du projet.

Valeurs :

- Rhapsodie

### Title

Titre complet du projet. «Rhapsodie. Corpus intonosyntaxique de référence en français parlé ».

### Contact Name

Nom de la personne responsable du projet.

### ContactAddress

Adresse de la personne responsable du projet.

### Contact.Email

Adresse email de la personne responsable du projet.

## Contact.Organisation

Nom de l'organisation responsable du projet.

## Description

Une description des objectifs du projet.

## Mediafile

Cette composante regroupe les informations sur les fichiers audio de chaque échantillon. Pour cette composante sont précisées les informations suivantes:

## ResourceLink

Lien vers le fichier audio.

## Type

Type du fichier multimédia.

Valeurs :

- audio

Interrogeable :NON

## Format

Le format du fichier multimédia.

Valeurs :

- audio
- x-wav
- mp3

Interrogeable :NON

## Size

Taille du fichier wav.

Valeurs :

- Megabytes
- Secondes

Interrogeable :NON

## Quality

Qualité du fichier audio

Valeurs :

- 0
- 1
- 2
- 3
- 4
- 5

Interrogeable : OUI

### **Recording**

Conditions description des conditions techniques concernant l'enregistrement du fichier multimédia.

### **Access publisher**

Nom du projet dans lequel le fichier apparaît.

### **Access Contact Name**

Nom de la personne responsable du projet (à contacter pour avoir accès au fichier)

### **Access ContactAddress**

Adresse de la personne responsable du projet (à contacter pour avoir accès au fichier)

### **Access Contact.Email**

Adresse email de la personne responsable du projet (à contacter pour avoir accès au fichier)

### **Access Contact.Organisation**

Nom de l'organisation responsable du projet (à contacter pour avoir accès au fichier)

### **Written resource**

Cette composante regroupe les informations concernant les annotations de chaque échantillon. Deux fiches sont créés l'une pour l'annotation syntaxique, l'autre pour l'annotation prosodique. Pour chaque type d'annotation sont précisées les informations suivantes :

### **ResourceLink**

Un lien au fichier contenant l'annotation

### **Type**

Type de ressource écrite.

Valeurs :

- annotation

Interrogeable :NON

### **Subtype**

Sous-type de ressource écrite.

Valeurs :

- Prosody : annotation prosodique
- Syntax : annotation syntaxique

Interrogeable :NON

## Format

Format du fichier utilisé pour la ressource écrite.

Valeurs :

Pour la syntaxe

- text/plain
- xml

Pour la prosodie

- textgrid
- xml

Interrogeable :NON

## Size

Taille de la ressource en termes de mots.

Valeurs :

- nombre de mots

## Validation.Type

Décrit la méthodologie et les outils utilisés pour valider l'annotation.

Valeurs :

- semi-automatic : l'annotation a été validée en utilisant des outils automatiques supervisés par des annotateurs experts
- manual: l'annotation a été validée manuellement

Une description textuelle précise les outils utilisés

## Character.Encoding

Nom de l'encodage des caractères utilisé dans les ressources écrites.

Valeurs :

- UTF-8

Interrogeable :NON

### **Content.Encoding**

Décrit les éléments annotés.

Valeurs :

Pour la syntaxe

- lemmatisation; catégorie grammaticale; liens de dépendance; fonctions syntaxiques; macrosyntaxe

Pour la prosodie

- proeminences syllabiques, dysfluences, contours globaux, périodes
- proeminences syllabiques, dysfluences, contours syllabiques contours globaux, périodes

Interrogeable :NON

### **Acces publisher**

Nom du projet dans lequel le fichier apparait.

Valeurs :

- Rhapsodie

### **Access Contact Name**

Nom de la personne responsable du projet (à contacter pour avoir accès au fichier)

### **Access ContactAddress**

Adresse de la personne responsable du projet (à contacter pour avoir accès au fichier).

### **Access Contact.Email**

Adresse email de la personne responsable du projet (à contacter pour avoir accès au fichier).

### **Access Contact.Organisation**

Nom de l'organisation responsable du projet (à contacter pour avoir accès au fichier).

## **Sources**

Cette composante regroupe les informations sur les projets source de chaque échantillon de Rhapsodie

## **ID**

Nom du projet source

Valeurs :

- PFC
- CFPP2000
- C-PROM
- ESLO
- Corpus Avanzi
- Corpus Lacheret
- Corpus Mertens
- Rhapsodie-PROF
- Rhapsodie-Movie
- Rhapsodie-BROAD

Interrogeables : OUI

## **AccessOwner**

Nom du responsable du projet source

Interrogeables : NON

## **Access.Contact.Organisation**

Nom de l'organisation responsable du projet source

Interrogeables : NON

## **Description**

Description textuelle et références bibliographiques du projet source

Interrogeables : NON