

Annexe 2.

ARC-FLUENCY

Set d'étiquettes des données et métadonnées

Crible Ludivine
ludivine.crible@uclouvain.be

Grosman Iulia
iulia.grosman@uclouvain.be

Dumont Amandine
amandine.dumont@uclouvain.be

Notarrigo Ingrid
ingrid.notarrigo@unamur.be

OBJECTIF : Etablir des traits communs pour les métadonnées du corpus, des situations et des locuteurs.

APPLICATIONS : Les noms d'étiquettes (colonne de gauche) et leur ordre donne une ligne de conduite lors de l'encodage des données par les doctorantes. L'uniformisation permettra un traitement technique des données plus automatique et fluide entre les différents corpus et enregistrements, situations et locuteurs étudiés au sein du projet ARC Fluency & Disfluency Markers.

METADONNEES DU CORPUS

COR.MTD	Métadonnées du corpus
COR.ID	nom d'identification du corpus
SUBCOR.ID	nom du ou des corpus sources
COR.AUTHOR	auteur du corpus
COR.CONTACT.AUTHOR	email de l'auteur du corpus
COR.DATE	date de création du corpus
COR.RIGHTS	droit d'utilisation/licence du corpus
COR.INSTIT	institution où le corpus a été complilé
COR.DURATION	total duration of the corpus (hh:mm:ss)
COR.NB.WORDS	total number of words in the corpus

METADONNEES D'UN TEXTE

TXT.MTD	Métadonnées du texte
TXT.ID	nom d'identification du texte
TXT.COR.ID	corpus d'origine du texte
DATE.RECORDING	date de l'enregistrement (YYYY-MM-DD)
TXT.URL	url direct vers le son ou la vidéo si existant
RIGHTS	droits d'auteurs/licence
TXT.DURATION	durée du fichier son (mm,ss)
TXT.NB.WORDS	nombre de mots dans la transcription
TXT.FILES	liste des fichiers associés avec le texte (.wav, .TextGrid, .exmeralda, .elan, etc.)

TEXT	
TEXTGRID/EXMARALDA/ELAN	
XML	
WAV/MP3	

TRAITS SITUATIONNELS

(1) Ce set cherche à fournir au chercheur les variables permettant une classification paramétrique des situations rencontrées dans le corpus oraux sans se référer à des labels pré-définis de tâches discursives.

(2) Ce set offre une liste de traits applicable à une vaste majorité de situations orales rencontrées dans une variété de corpus.

(3) Ce set a pour but de fournir un cadre théorique d'analyse de l'influence des traits situationnels entre eux, i.e. les valeurs des traits situationnels s'envisagent comme faisant partie d'un profil situationnel complexe et non comme des variables isolées. En vue d'opérationnaliser l'analyse en traits situationnels, le nombre de ces derniers a volontairement été réduit. La présente liste inclu ainsi 6 traits situationnels ayant été admis comme variables centrales de toute situation orale. Les exemples fournis correspondent à la forme prototypique de la situation citée.

(4) Afin de parvenir à une catégorisation plus fines et complète, l'utilisation de ce set n'interdit en rien le chercheur d'y ajouter des traits situationnels complémentaires tels la relation entre locuteurs, la connaissance du sujet, le type de diffusion médiatique concerné, etc.

(5) Le protocole d'annotation envisage :

- Les traits situationnels comme s'appliquant à une situation et non aux locuteurs de la situation
- Tous les traits définis graduellement, à l'exception de celui se référant au "nombre d'interlocuteurs"

SIT.FEATURE	(Traits situationnels)
TXT.ID	Nom d'identification du texte
D.ELICIT	Degré d'élicitation
	La présence et le poids du protocole expérimental agissant comme une contrainte sur l'interaction
DIRIGE	Production artificielle due à un protocole expérimental rigide dans le cadre d'une recherche scientifique exerçant un contrôle fort sur la production du locuteur et sur des variables linguistiques.
SEMI.DIRIGE	Production naturelle supervisée par un protocole expérimental souple dans le cadre d'une recherche scientifique guidant le choix du topique mais ne contraignant pas le locuteur à choisir une formulation préalable. Le locuteur est libre du choix de ses mots.
NATUREL	Production authentique du locuteur dénuée de tout protocole expérimental, non-suscitée par une action de recherche scientifique.
NB.SPK	Catégorisation du nombre de locuteurs
	Le nombre de locuteurs prenant une part active dans l'interaction
MONOLOGUE	Une seule personne est le locuteur principal désigné.
DIALOGUE	Deux personnes sont les locuteurs principaux désignés

MULTILOGUE	Plus de deux locuteurs principaux prennent part à l'interaction.
D.PREP	Degré de préparation
	Le degré de la préparation orale et/ou écrite du discours du locuteur principal
PREPARE	Le locuteur a entièrement préparé avant l'interaction le contenu et la forme de son discours. Le discours a été scripté et est formulé avec ou sans support textuel.
SEMI.PREPARE	Le locuteur a préparé la direction globale de son discours, mais le discours n'a pas été scripté dans sa totalité. L'interaction peut inclure un support textuel partiel.
SPONTANE	Le locuteur improvise son discours au fur-et-à-mesure de l'interaction et n'en a préparé aucune partie.
D.INTERACT	Degré d'interactivité
	La capacité d'adaptation des interlocuteurs au regard des attentes sur le statut des locuteurs dans l'interaction.
INTERACTIF	Relation symétrique entre les locuteurs. La situation en cours permet aux interlocuteurs de prendre la parole.
SEMI.INTERACTIF	Relation asymétrique où un locuteur occupe plus la parole qu'un autre. La situation de parole n'exclut pas des interventions ponctuelles des locuteurs secondaires.
NON.INTERACTIF	Relation asymétrique où un des locuteurs prend et garde de manière quasi-continue la parole, sans possibilité d'intervention des interlocuteurs.
D.MEDIA	Degré de médiatisation
	L'importance de la diffusion sur les grands média comme objectif principal de l'interaction
MEDIATISE	Le but principal de l'interaction est d'être diffusé par un média.
SEMI.MEDIATISE	L'interaction aurait lieu même si elle n'était pas diffusée par un média.
NON.MEDIATISE	L'interaction n'est diffusée par aucun média.
C.PRO	Catégorie de professionnalisation
	La situation est ou n'est pas suscitée par la profession des interlocuteurs
PRO	La situation est suscitée par la profession d'au moins un des interlocuteurs.
NON.PRO	La situation n'est pas suscitée par la profession des interlocuteurs.

METADONNES DES LOCUTEURS

SPK.MTD	
SPK.CODE	Code d'identification du locuteur
SPK.TXT.ID	texte(s) dans le(s)quel(s) le locuteur intervient
AGE.N	âge du locuteur lors de l'enregistrement
SPK.AGE	Tranches d'âge du locuteurs lors de l'enregistrement
18-25	Le locuteur appartient au groupe d'âge entre 16 et 25 ans
26-45	Le locuteur appartient au groupe d'âge entre 26 et 45 ans
46-65	Le locuteur appartient au groupe d'âge entre 46 et 65 ans
66+	Le locuteur appartient au groupe d'âge de 66 et plus
SPK.GENDER	Sexe du locuteur
F	FEMININ
M	Masculin
SPK.NATIONALITY	Nationalité. Pays d'origine du locuteur (en toute lettre)
SPK.1LG	Langue maternelle du locuteur
SPK.OTHERLG	Autre(s) langue(s) maîtrisée(s) par le locuteur
SPK.JOB	La relation du locuteur avec les métiers de la communication
COM	Le locuteur occupe une fonction dans les métiers de l'information et/ou de la communication.
NON-COM	Le locuteur ne travaille pas dans les métiers de l'information et/ou de la communication.

...A venir

Set d'étiquettes des marques de disfluece locales (?)

Références

Koch P. & Osterreicher W. 2001. « Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit ». In G. Holtus, M. Metzeltin & C. Schmitt (eds), Lexikon der Romanistischen Linguistik. Bd. I / 2. Tübingen, Niemeyer. 584-627.

Goldman J.-Ph., Auchlin, A., & Simon, A. C. 2011. "Discrimination de styles de parole par analyse prosodique semi-automatique". In H.-Y. Yoo & E. Delais-Roussarie (eds), Actes d'IDP 2009, Paris, 9-11 septembre 2009 (p. 287 301). Présenté à IDP 2009, Paris.

(suite des références, à venir).

REF NEEDED : Parole professionnelle non professionnelle
Tranche d'âge
DELIVERY
BDU
POS/RECTION ETC.
INTERACTIVITE
MEDIATISATION & PAROLE

Annexes (brouillon en reste)

ANNOTATION.TIER	couches d'annotation associées avec la situation
TIER.NAME	(add your suggestions)
ORTHO	transcription orthographique
WORD	transcription en mot
SYLL	segmentation syllabique
DELIVERY	transcription des phénomènes paralinguistiques
PHON	segmentation phonétique
PROM	annotation en proéminence
ACCENT	annotation accentuelle
CLITIC	annotation clitique/non clitique
POS	annotation en partie du discours
RECTION	annotation en unités rectionnelles
BDU	annotation en Basic Discours Units
MD	annotation des marqueurs de discours
DIS	annotation des disfluences

TIER.MD	Métadonnées de la tire
UPDATE	Création/Dernière révision
ANNOT.METHOD	Méthode d'annotation
AUTO	méthode automatique (nom du logiciel)
MANU	méthode manuelle (auteur)
BIB.REF	Support Référence bibliographiques
DEPENDANCE	Tier sur lequel se fonde cette annotation
COMMENT	Champ libre (tâche à réaliser)
	(add your suggestions)