Projet AADI : Aphasie et Analyse du Discours en Interactions : constitution de bases de données et nouvelles méthodes d'exploitation. Manuel d'instructions : méthodologie de recueil, transcription et codage des données

Halima Sahraoui, Silvia Martínez-Ferreiro, Aleksandra Nowakowska

Projet AADI

Aphasie et **A**nalyse du **D**iscours en **I**nteractions : constitution de bases de données et nouvelles méthodes d'exploitation

Manuel d'instructions : méthodologie de recueil, transcription et codage des données

Sahraoui, H., Martínez-Ferreiro, S. & Nowakowska, A.

Projet Région Occitanie – FEDER : N°2019-A03105-52

PROJET COFINANCÉ PAR LE FONDS EUROPÉEN DE DÉVELOPPEMENT RÉGIONAL

0 Cadre général

L'objectif principal de du projet AADI est de collecter, de transcrire et annoter des données de discours et interactions recueillies auprès de personnes aphasiques (N=50) et contrôles (N=60), afin de constituer une base de données en langue française accessible *via* Ortolang à exploiter selon diverses perspectives de recherches linguistiques et cliniques.

0.1 Références du projet

Références du projet :

Projet AADI - Aphasie et Analyse du Discours en Interactions : constitution de bases de données et nouvelles méthodes d'exploitation (2018-2022).

Projet Région – FEDER (financement Région Occitanie et Fond Européen de Développement Régional) N°2019-A03105-52.

Laboratoire porteur :

- Aleksandra Nowakowska, Praxiling UMR 5267 & Université Paul Valery Montpellier 3

https://praxiling.cnrs.fr/presentation-2/

Laboratoires partenaires :

- Halima Sahraoui, Barbara Köpke, Laboratoire de NeuroPsychoLinguistique LNPL - E.A. 4156, Université de Toulouse

https://lnpl.univ-tlse2.fr/

- M. Barkat- Defradas, Institut des Sciences de l'Évolution de Montpellier ISEM – UMR 5554

https://isem-evolution.fr/

Entreprise partenaire :

- Lionel Fontan, Archean Labs – Société Archean Technologies, Montauban

https://www.archean.tech/archean-labs-en.html

Personnel ingénieur d'études ou de recherches sous contrat :

- LNPL Toulouse : Silvia Martínez-Ferreiro
- PRAXILING Montpellier : Dodji Gbedahou, Typhanie Prince, Ode Dor

0.2 Comité d'éthique et fiche de traitement

Un avis favorable a été rendu par le Comité d'Éthique de l'Université Fédérale de Toulouse (11 décembre 2020, Projet 2020-268). Une fiche de traitement a été déposée auprès du Délégué à la Protection des Données de l'Université Paul Valéry - Montpellier 3.

1 Instructions de passation

1.1 Architecture de passation : participants aphasiques et contrôles

Le protocole AADI est conçu pour être administré en 1 à 3 sessions avec des participants aphasiques et en 1 à 2 sessions avec des participants contrôles (sans la passation des pré-tests cliniques).

Session	Tâche (1, 2, 3, 4, 5)	
Pré-recueil		Premier contact / inclusion Notice d'information et consentement éclairé (oral et écrit) Questionnaire et données démographiques Passation des tests d'évaluation (seulement pour les participants aphasiques
Recueil 01	**1** **2** **3**	Entretien orienté par des questions Description à partir d'une image Narration d'histoire
Recueil 02	**4** **5**	Lecture de phrases Entretien, communication authentique (situation écologique)

Le recueil de ces données orales donne lieu à une transcription sous le logiciel CLAN et selon les standards de codages CHAT (Mac Whinney, 2021 ; Bernstein Ratner, Brundage & Fromm, 2020, cf. section 2).

Durée de recueil : La durée estimée de recueil de données (y compris les durées estimées des instructions et pauses pendant les passations) est de 1h30 à 2h pour les participants aphasiques. Le nombre de sessions de recueil dépendra du participant aphasique, car la durée de chaque session devra toujours respecter l'état du patient (fatigue, sommeil, …). Quel que soit le choix du participant aphasique, et en fonction des difficultés relevées en cours de passation, l'examinateur lui proposera de poursuivre le recueil dans une session ultérieure.

1.2 Préparation matérielle

En présentiel : idéalement, réaliser le recueil dans une pièce au calme, à l'aide d'un enregistreur audio numérique, et une caméra sur pied. Un ordinateur portable équipé (caméra, micro de bonne qualité) peut aussi être utilisé.

En distanciel : en raison de la situation d'urgence COVID-19 et en fonction de l'évolution de la situation sanitaire, les participants qui le souhaitent pourront faire la passation en ligne selon les moyens techniques à disposition et sans obligation.

1.3 Premier contact / inclusion : participants aphasiques

Critères généraux d'inclusion :

- Être âgé de 18 ans minimum et volontaire
- Consentement à la participation à l'étude
- Participants diagnostiqués aphasiques (dont Aphasie Progressive Primaire) (diagnostic sur des bases cliniques, comportementales et/ou neuroimagerie)
- Lésion cérébrale uni - ou bilatérale
- Stade aiguë, chronique, administration du protocole à compter d'un mois de la survenue de l'aphasie
- Pas d'antécédents de désordre d'origine neurologique autre que ce qui a causé l'aphasie (antécédents de troubles psychiatriques, troubles développementaux, maladie neurodégénérative sauf pour les APP)
- Vision et audition normales ou corrigées
- Capacité à comprendre et suivre les instructions du protocole sur la base des explications adaptées au type et sévérité de l'aphasie (les instructions sont adaptées pour les participants aphasiques avec trouble de la compréhension)

Critères généraux de non inclusion :

- Aphasie infantile
- Accidents Ischémiques Transitoires
- Antécédents de désordre d'origine neurologique (troubles psychiatriques, troubles développementaux, maladie neurodégénérative sauf pour les APP, etc…)
- Présenter des troubles visuels ou auditifs non corrigés
- Incapacité à comprendre ou suivre les instructions du protocole
- (l'apraxie et la dysarthrie ne sont pas des critères d'exclusion ; niveau d'études ou de littéracie non plus)

Critères supplémentaires d'inclusion COVID-19 pour les passations en présentiel :

- Selon les recommandation de l'ARS en vigueur
- En phase chronique à partir de 3 mois post-AVC avec pathologie stabilisée
- < 65 ans
- Sans HTA non contrôlée
- Sans obésité
- Sans insuffisance respiratoire
- Sans traitement immuno-dépresseur

1.4 Premier contact / inclusion : participants contrôles

Critères généraux d'inclusion :

- Être âgé de 18 ans minimum et volontaire
- Consentement à la participation à l'étude
- Pas d'antécédents de désordre d'origine neurologique (antécédents de troubles psychiatriques, troubles développementaux, maladie neurodégénérative)
- Vision et audition normales ou corrigées
- Capacité à comprendre et suivre les instructions du protocole sur la base des explications.

Critères généraux de non inclusion :

- Antécédents de désordre d'origine neurologique (troubles psychiatriques, troubles développementaux, maladie neurodégénérative, etc…)
- Présenter des troubles visuels ou auditifs non corrigés
- Incapacité à comprendre ou suivre les instructions du protocole

Critères supplémentaires d'inclusion COVID-19 pour les passations en présentiel : selon les recommandation de l'ARS en vigueur (non inclusion des personnes atteints d'une maladie chronique ou fragilisant leur système immunitaire : antécédents cardiovasculaires, diabète et obésité, pathologies chroniques respiratoires, cancers, insuffisance rénale, …).

1.5 Lettre d'information et consentement éclairé

Avant les passations du protocole, les participants doivent avoir pris connaissance de toutes les informations fournies dans la lettre d'information et avoir signé le formulaire de consentement éclairé. En matière de règlementation (et son évolution récente) dans les domaines de recherche nécessitant le recueil de corpus en pathologies du langage, lire notamment Lalain *et al.* (2021). Le code d'anonymat à renseigner dans le formulaire de consentement éclairé est le suivant : un numéro aléatoire à 2 chiffres suivi des chiffres et lettres associées aux dates de recueil : 83-2021-03-01 (pour un recueil fait avec le participant 83 du 01 mars 2021).

1.6 Questionnaire et données démographiques

Le questionnaire a été formalisé selon le protocole de la base de données *Aphasiabank* à partir de l'adaptation au français réalisée par Colin & Le Meur (2016, https://aphasia.talkbank.org/, MacWhinney *et al.*, 2011), qui a encore été ajusté par nos soins pour le protocole AADI. Le questionnaire « données démographiques » est complété à partir des informations mises à disposition par le participant (documents divers) et avec l'assistance de l'examinateur sous la forme d'un entretien complémentaire si besoin.

1.7 Passation des tests : participants aphasiques

Les tests suivants ont été sélectionnés suivant l'adaptation française du protocole Aphasiabank réalisée par Colin & Le Meur (2016).

I. Test de dénomination de noms MT86 (Lecours et al., 1996) ;

II. Test de dénomination de verbes DVL-38 (Hammelrath, 1999; Hammelrath et al., 2000) ;

III. Test de compréhension orale de mots MT86 (Lecours et al., 1996) ;

IV. Test de compréhension orale de phrases (20 premiers items) MT86 (Lecours et al., 1996) ;

V. Échelle de gravité de l'aphasie BDAE - HDAE (Goodglass et al., 2001a, 2001b; Goodglass & Kaplan, 1972; Mazaux & Orgogozo, 1982 ; voir page 3 du livret). Pour ce test en particulier, il convient d'apprécier la gravité de l'aphasie d'après la première tâche de langage spontané du protocole AADI, donc en cours de passation du recueil 01.

Ces tests ne sont pas supposés être administrés s'il y a déjà eu une évaluation avec ces tests dans les 6 mois précédents pour un participant aphasique en phase chronique, et dans les 2 mois précédents pour un patient aphasique en phase aiguë.

Cependant, il est toujours mieux de faire passer ces tests ciblés et standardisés et ainsi harmoniser les résultats de tests objectivés complémentaires aux données discursives de la base de données AADI.

Le temps estimé de passation si elle est nécessaire : 30 minutes à une heure selon le profil du participant aphasique.

Remarque : il est nécessaire de coter les réponses à ces tests dans le tableur dédié « AADI Pré-tests Aphasiques ». Ce tableur récapitulatif se complète automatiquement au fur et à mesure que les réponses sont cotées dans les onglets. Il convient de compléter un fichier par participant aphasique, en utilisant le même code anonymat pour nommer le fichier.

1.8 Instructions générales

Concernant les tâches de discours 1, 2 et 3, les instructions suivantes exposent la trame que doit suivre l'examinateur. Un script à suivre le plus fidèlement possible a été fourni aux examinateurs. L'examinateur doit lire toutes les instructions avant de procéder au recueil.

NB : Le script détaillé guidant les interventions de l'examinateur n'est pas fourni dans ce manuel. Les scripts de passation peuvent être fournis sur demande aux laboratoires partenaires du projet.

1) Laisser suffisamment de temps pour que le participant puisse compléter sa réponse, tout en respectant le script du protocole.

2) Selon les consignes et pour faciliter la transcription, les paroles de l'examinateur, y compris les encouragements verbaux, doivent être restreints au strict nécessaire. Utiliser les **encouragements non verbaux** (hochements de tête, expressions faciales, contact visuel) plutôt que verbaux (« Je vois », « hum hum », « euh », « oui », « bien sûr », « je crois aussi », « très bien », etc.), de façon à éviter d'orienter la production du participant.

3) Il faut également, dans la mesure du possible, ne pas interrompre le participant lorsque celui-ci est en train de parler, lui laisser tout le temps de poursuivre son propos.

Questionnaire d'aide à utiliser en cas d'échec aux questions classiques du protocole :

Certains participants ne seront pas capables de répondre aux questions du protocole, même avec les incitations fournies dans le script. À la fin de chaque script, nous avons inclus un "Questionnaire d'aide". Ceci constitue un supplément destiné à permettre à ceux qui ont une atteinte langagière plus sévère de répondre aux épreuves. Pour chaque question, une ou plusieurs questions supplémentaires sont proposées afin d'aider le participant à produire une réponse. Ce questionnaire d'aide permet ainsi à l'examinateur de relancer ou de terminer l'entretien de façon plus simple si la personne aphasique a trop de difficulté pour continuer.

Il est possible de fournir une carte imagée avec OUI/NON au participant si l'atteinte langagière est trop sévère pour qu'il puisse répondre. Si, malgré les sollicitations

supplémentaires, le participant n'est pas capable de répondre, l'examinateur met fin à la passation.

Pour toutes les tâches de discours, il est recommandé d'obtenir au minimum 700 mots (Ossewaarde *et al.*, 2020).

a. Recueil 01

Commencer par une conversation préalable, non enregistrée, pour effectuer un premier échange avec le participant. Expliquer et faire signer le formulaire de consentement, expliquer que la passation est filmée / enregistrée, répondre à toutes les questions du participant.

COMMENCER L'ENREGISTREMENT

Important : enregistrer le préambule suivant sur le consentement éclairé, puis débutez la passation de la première tâche du protocole (histoire de l'aphasie et récupération).

Examinateur : « Le / La participant(e) a signé un formulaire de consentement éclairé. Ces données ne peuvent être utilisées que pour la recherche et l'enseignement ».

- **Tâche 1 : Entretien orienté par des questions**

Objectif : recueillir du discours autobiographique en orientant la production avec des questions :

a) Histoire de l'aphasie (ou une maladie pour les contrôles) et récupération

b) Évènement important

c) Dernières vacances

Temps estimé de passation et nb de mots :

- Participants aphasiques fluents : 7 mn
- Participants aphasiques non fluents : 17 mn
- Participants contrôles : 7 mn
- Nb de mots attendu : 700

Si les participants ne sont pas capables de répondre aux questions ouvertes, utiliser le « questionnaire d'aide », autant que nécessaire. Si les participants sont incapables de répondre aux questions d'aide de la partie A (histoire de l'aphasie et réadaptation) de la tâche 1 (échantillon de discours libre), ne pas réaliser la partie b. (événement important), ni c. (dernières vacances) et poursuivre en passant directement à la tâche 2 (description d'image).

- **Tâche 2 : Description à partir d'une image**

Objectif : à partir d'un dessin épuré en noir et blanc (*Le Sauvetage du Chat*, *Aphasiabank*), les participants sont invités à raconter une histoire avec un début, un milieu et une fin pour obtenir une narration. Support imagé : https://aphasia.talkbank.org/protocol/english/participant/singleStimPages/cat.html

Temps estimé de passation et nb de mots :

- Participants aphasiques fluents : 7 mn
- Participants aphasiques non fluents : 17 mn
- Participants contrôles : 7 mn
- Nb de mots attendu : 700

NB : selon les mêmes instructions, une deuxième image d'une scène de la vie quotidienne (Scène de Pique-Nique, WAB-R) sera proposée si cela est nécessaire pour amener à la production de nombre de mots attendu (700 mots minimum).

- **Tâche 3 : Narration d'histoire**

Objectif : raconter l'histoire de *Cendrillon* de mémoire. Si le participant ne parvient pas à se remémorer l'histoire, il peut s'appuyer sur un livre d'images ou un diaporama sur écran. Support imagé : https://aphasia.talkbank.org/protocol/english/participant/sequences/Cinderella/index.html

Temps estimé de passation et nb de mots :

- Participants aphasiques fluents : 7 mn
- Participants aphasiques non fluents : 17 mn
- Participants contrôles : 7 mn
- Nb de mots attendu : 700

b. Recueil 02

- **Tâche 4 : Lecture de 10 phrases**

Objectif : les participants lisent une série de 10 phrases plus ou moins complexes issues du BDAE-HDAE (Goodglass et al., 2001a; Mazaux & Orgogozo, 1982).

Temps estimé de passation :

- Participants aphasiques fluents : 1 à 2 mn
- Participants aphasiques non fluents : 5 mn
- Participants contrôles : 1 à 2 mn

Laisser tout le temps nécessaire au participant pour les lire. En cas d'impossibilité pour les 3 premières phrases ou au bout de 20 secondes sans réponse, aider le participant en donnant l'amorce du premier mot pour chaque phrase. En cas de très grandes difficultés, proposer au patient d'arrêter la tâche de lecture s'il le souhaite. Ne pas rester sur un « échec » en cas de grandes difficultés, proposer une lecture accompagnée (l'examinateur lit, le participant répète).

La tâche de lecture de 10 phrases peut être fatigante, même si elle n'est pas achevée. Laisser un temps de pause suffisant avant de passer à la tâche 5.

- **Tâche 5 : Entretien, communication authentique (situation écologique)**

Rappel : tous les participants doivent avoir signé un formulaire de consentement éclairé avant de commencer l'enregistrement.

Objectif : recueillir des interactions en situation de communication plus authentique et à but relationnel selon un cadre situationnel comptant 2 interlocuteurs.

Thème A : « Si on parlait du temps qu'il fait ? »

Thème B : « Si vous gagnez 10 000 000 d'euros au loto, que feriez-vous ? »

Thème C : « De quoi souhaitez-vous parlez ? »

Temps de recueil : 20 minutes minimum (durées moyenne des interactions analysées dans les ouvrages spécialisés, cf. Kerbrat-Orecchioni, 2013).

-Nombre de participants à l'interaction : 2 (dyade : participant aphasique et personne proche ou ami ou famille ou aidant régulier) ; 2 personnes proches (dyade : participant contrôle ami ou famille).

-Lieux et moments : au choix des participants, selon les situations et selon la convenance des participants afin de s'adapter à eux et éviter la fatigabilité des participants aphasiques en particulier (lieu calme et propice à une communication non dégradée par l'environnement).

Complément optionnel tâche 5 : interactions dans des situations variées non prédéfinies : recueil complémentaire et optionnel si l'opportunité se présente et au choix des participants, dans des situations de communication variées, les plus naturalistes possibles. Le choix des participants, des lieux et du moment de recueil incombent aux participants.

2 Transcription et codages de base sous CLAN

2.1 Avant de commencer

1. Préparer les fichiers média audio / vidéo : s'il y en a plusieurs, concaténer les fichiers audio / vidéo pour ne créer qu'un seul fichier au final comportant, les uns à la suite des autres, tous les recueils.

2. Installation de CLAN : suivre les instructions du manuel / tutoriel Clin-CLAN :

Manuel très synthétique et plus simple auquel il convient de se référer :

- Bernstein Ratner, N., Brundage, S.B. & Fromm, D. (2020), Clin-CLAN - A Clinician's Complete Guide to CLAN and PRAAT, updates by D. Fromm : https://talkbank.org/manuals/Clin-CLAN.pdf.

Manuels complets originaux :

- MacWhinney, B. (2000 [2022]). *The CHILDES Project: Tools for Analyzing Talk*. *Part 1: The CHAT Transcription Format.* 3 rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates [on-line updated 2022 version : https://www.talkbank.org/manuals/CHAT.pdf].

- MacWhinney, B. (2000 [2022]). *The CHILDES Project: Tools for Analyzing Talk*. *Part 2: The CLAN Program.* 3 rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates [on-line updated 2022 version : https://www.talkbank.org/manuals/CLAN.pdf]

Télécharger CLAN & la grammaire MOR : https://talkbank.org/ (voir manuel et tutoriels) :

-Télécharger le logiciel CLAN ici : https://talkbank.org/ dans « Programs »

-Ouvrir CLAN > Aller dans Fichier > Télécharger la grammaire MOR > Français et l'installer (MOR grammar servira notamment à étiqueter morphologiquement le corpus par la suite, après avoir transcrit les données sous la forme de fichiers .cha conformes).

2.2 Création Fichiers .cha : anonymat et nom de fichier

2.2.1 Code anonymat

CODE ANONYMAT : un numéro aléatoire à 2 chiffres suivi des chiffres et lettres associées aux dates de recueil : 83-2020-03-01 (pour un recueil fait avec le participant 83 du 01 mars 2020).

Pour pouvoir détruire les fichiers sans avoir recours à des données personnelles non conservées (nom), un code anonymat est créé pour chaque participant, et donné dans son formulaire de consentement (si le participant nous contacte pour supprimer ses données recueillies, il n'a qu'à communiquer ce code). Ce code doit donc apparaître dans le nom du fichier.

2.2.2 Modèle : nom de fichier .cha

Mettre toujours au début « AADI-ORI », puis le nom de l'investigateur principal responsable de la collecte et suivi des transcriptions, puis « aph » ou « contr » (participant aphasique ou contrôle) selon le modèle ci-dessous :

Adopter ce format de nom de fichiers :

AADI-ORI_nom investigateur_aph_code anonymat.cha
AADI-ORI_nom investigateur_contr_code anonymat.cha

Exemples :
AADI-ORI_sahraoui_aph_83-2020-03-01.cha
AADI-ORI_martinez_contr_32-2021-04-06.cha

AADI-ORI signifie « original ». Il s'agit ainsi de garder une transcription AADI-ORI comme transcription brute originale (sans lignes %MOR), avec les codages de base que nous listons dans la suite.

2.2.3 Modèle : nom de fichier audio ou vidéo et alignement F5

Le nom de fichier audio ou vidéo doit être le même que le nom de fichier .cha.

2.3 Sauvegarde des fichiers .cha et audio-vidéo

1 fichier .cha correspond à 1 participant et à 1 fichier vidéo/audio
1 participant = 1 seul fichier .cha = 1 seul fichier vidéo/audio

La sauvegarde des fichiers est réalisée conformément aux instructions de la fiche de traitement.

2.4 Transcription et codage

Afin de transcrire et coder les échantillons de discours spontané, suivre les étapes suivantes :

1) Création suivant la structure du fichier .cha, headers, fichiers média associés
2) Transcription orthographique des lignes principales, avec les codages de base
3) Vérification des erreurs ESC-L : conformité du fichier .cha
4) Alignement F5 avec le fichier média

2.4.1 Structure du fichier .cha (squelette – template), headers, fichier média

a. Structure globale du fichier .cha

La structure du fichier .cha correspond à l'enchainement des recueils 01 et 02, tâches 1, 2, 3, 4, 5. Dans le tableau ci-dessous, les tâches du recueil sont rappelées à gauche, et la notation correspondante dans le fichier .cha à droite. Lorsqu'on passe d'une tâche à l'autre, cela est spécifié avec la notation @G:

TÂCHES PROTOCOLE AADI	Fichier .CHA - TEMPLATE CLAN @GEM
Tâche 1 - Entretien orienté par des questions : a) Histoire de l'AVC et récupération / histoire d'une maladie et récupération b) Évènement important c) Dernières vacances.	@G: Stroke @Time Duration: 00:07:27 *INV: bonjour . *PAR: bonjour . @G: Important_Event @Time Duration: 00:07:17 @G: Holidays @Time Duration: 00:07:17
Tâche 2 - Description à partir d'une image : Le sauvetage du chat (et une 2ème description si nécessaire avec l'image du Pique-Nique – Wab-R)	@G: Cat @Time Duration: 00:07:16 @G: Second_Picture @Time Duration: 00:07:16
Tâche 3 - Narration d'histoire : Cendrillon Sinon : -Petit Chaperon Rouge (Little Red Riding Hood) -Autre conte	@G: Cinderella @Time Duration: 00:07:36 @Comment: LRR @Time Duration: 00:07:36 @Comment: Other_Fairy_Tale @Time Duration: 00:07:36
Tâche 4 - Lecture de phrases :	@G: Sentences @Time Duration: 00:02:44
Tâche 5 - Entretien, communication authentique (situation écologique) : Thème de discussion proposé aux participants pour initier et maintenir la conversation (Dialogue Aph / Contr ou Contr / Contr) Complément Tâche 5 : Interactions dans des situations variées non prédéfinies	@G: Conversation_Dialog @Time Duration: 00:20:44 @Comment: @G: Interactions_Complementary @Situation: @Activities: @Comment: @Time Duration: 00:20:44 @End

Les @Gs *Stroke, Important_Event, Cat, Cinderella* sont harmonisés avec les repères de transcriptions d'*Aphasiabank*, ce qui permet des analyses automatiques sur une plus large base (avec des corpus francophones ou d'une autre langue existants).

a. Modèle de fichier .cha à adopter strictement

Un fichier de base (ou *template*) doit respecter les formats ci-après. Un exemple de fichier conforme transcrit est fourni à la fin de ce manuel (section 4.).

Le **nom de ce fichier de base : AADI-ORI_nomINV_aphcontr_CodeAnonymat.cha**

La structure .cha du template est la suivante :

```
@Begin
@Languages:	fra
@Participants:	INV Investigator, PAR Participant, PAR1 Partner
@ID: fra|AADI_Toulouse|INV|40;05.00|female|||Investigator|||
@ID: fra|AADI_Toulouse|PAR|65;03.10|male|aphasia|ANO|Participant|||
@ID: fra|AADI_Toulouse|PAR1|50;06.04|female|||Partner|||
@Transcriber:	student_name
@Transcriber:	supervisor_martinez-ferreiro_sahraoui
@Warning:
@Media:	AADI-ORI_nomINV_aphcontr_CodeAnonymat, video
@Date:	26-JUN-2012
@Time Duration:	01:10:40
@G: Stroke
@Time Duration:	00:10:27
*PAR:	bonjour SA boujour à toutes et à tous . •0_1328•
*INV: alors vous avez l'air plus sereine et détendue ce matin . •1328_3340•
*INV: que se passe-t-il ? •3340_4152•
*PAR:	eh bien parce que depuis jeudi j'écoute ça . •4152_7367•
%com:	une musique passe
*INV: www . •7367_56809•
@G: Important_Event
@Time Duration:	00:07:32
@G: Holidays
@Time Duration:	00:06:17
@G: Cat
@Time Duration:	00:06:35
@G: Second_Picture
@Time Duration:	00:00:00
@G: Cinderella
@Time Duration:	00:08:36
@Comment:	LRRH
@Time Duration:	00:00:00
@Comment:	Other_Fairy_Tale
@Time Duration:	00:00:00
@G: Sentences
@Time Duration:	00:02:00
@G: Conversation_Dialog
@Time Duration:	00:11:00
@Comment:
@G: Interactions_Complementary
@Situation:
@Activities:
@Comment:
@Time Duration:	00:00:00
@End
```

a. Headers

En fonction des participants, il faudra adapter les identifiants des locuteurs et leur qualité, éventuellement ajouter des locuteurs. Par défaut : utiliser INV, PAR = participant cible aphasique ou contrôle, PAR1 = participant en plus, et continuer PAR2, PAR3, etc… selon la situation d'interaction en utilisant le menu dans «TIERS > ID Headers » pour les spécifier.

```
@Begin
@Languages:  fra
@Participants:  INV Investigator, PAR Participant, PAR1 Partner
@ID: fra|AADI_Toulouse|INV|40;05.00|female|||Investigator|||
@ID: fra|AADI_Toulouse|PAR|65;03.10|male|aphasia||Participant|||
@ID: fra|AADI_Toulouse|PAR1|50;06.04|female|||Partner|||
@Transcriber:  student_name
@Transcriber:  supervisor_name
@Warning:
@Media:        AADI-ORI_nomINV_aphcontr_CodeAnonymat, video
@Date:         26-JUN-2012
@Time Duration:        01:10:40

@Comment:

@Situation:
@Activities:
@Comment:

@End
```

Modifier les informations suivant le participant (par exemple l'âge, le type d'aphasie…). Pour le type d'aphasie, suivre ce système de classification de Boston, à noter dans l'en-tête dédiée au participant :

-pour un participant aphasique : @ID: fra|AADI_Toulouse|PAR|65;03.10|male|aphasia|BRO|Participant|||

-pour un participant contrôle : @ID: fra|AADI_Toulouse|PAR|65;03.10|male|control||Participant|||

NA Non Applicable
ANO Anomique
BRO Broca
CON Conduction
GLO Global
MTC Transcortical mixte
TCM Moteur transcortical
TCS Transcortical Sensoriel
WER Wernicke
NCL non classifiable
OTH autre
U indisponible

b. Conformité de transcription : Esc-L

Respecter rigoureusement les espaces, lignes, signes, formats de dates, etc.
Faire ESC-L : la mention « Success ! No errors found. » doit apparaître en bas de l'écran fichier, ce qui permet de savoir si le fichier est conforme. Si une erreur est signaler, corriger en se référant aux instructions des manuels CLAN / CHAT.

c. Fichier média et fichier .cha : nom et dossier commun

10 @Media: AADI-ORI_sahraoui_aph_83-2020-03-01, video
11 @Date: 26-JUN-2012
12 @Time Duration: 00:3:52

Le nom du fichier .cha doit être identique au nom du fichier média vidéo ou audio, et être placé dans le même dossier, ce qui permet la liaison / alignement. Aussi, si le fichier média est vidéo, il faut préciser « video », s'il est audio, il faut préciser « audio ».

2.4.2 Transcription des lignes principales

Les lignes (« Tier ») commençant par * indiquent ce qui a été réellement dit. Celles-ci sont appelées « lignes principales ». Exemple :

*INV: bonjour madame FE . ●

*PAR: bonjour monsieur XR . ●

La transcription est orthographique, mais des codages sont à ajouter (voir sections suivantes).

Chaque ligne principale doit inclure un seul énoncé. Lorsqu'un participant produit plusieurs énoncés dans une intervention, transcrire chaque énoncé sur une nouvelle ligne principale. Il s'agira donc de segmenter le discours en énoncés.

D'après Bernstein Ratner, Brundage & Fromm (manuel Clin-CLAN, 2020), un énoncé est une chaîne de mots qui :

1. est suivi d'une pause (même brève) ;
2. se termine par un contour d'intonation terminal ;
3. a une structure grammaticale complète.

2.4.3 Segmentation du discours continu en unités

La segmentation du discours oral en sous-unités de discours, ce que nous appelons énoncés (*utterances*) dans le cadre de ce recueil, est une étape importante car elle permet de fournir par exemple des mesures traduisant la longueur moyenne d'un énoncé ou le degré d'élaboration syntaxique.

Le discours est continu, mais il faut retourner à la ligne à chaque nouvel énoncé produit, et cela nécessite de définir des critères de segmentation à partir de l'oralité (alors qu'à l'écrit, on se réfère aisément aux points finaux des phrases). En général, on se réfère aux contours intonatifs pour segmenter. Cependant, certains locuteurs ont un débit verbal rapide et il peut être difficile de trouver « où » segmenter le flux pour aller à la ligne et isoler les unités de discours. Par ailleurs, il y a des situations où le locuteur fait une pause longue mais continue sa phrase ou va reformuler après 1 ou même plus d'un seconde de pause silencieuse.

Comme il s'agit de discours continu, les repères tels que les connecteurs de discours (particules d'initiation ou de clôture d'énoncés : *en fait, et puis, puis, alors, en fait, et alors, après, enfin, donc, etc.*) peuvent être de bons indices de frontières pour « segmenter » les unités discursives en unités de discours connectées entre elles.

À l'instar de Menn et Obler (1990 : 1377), on peut opérer une distinction entre (1) les particules conjonctives optionnelles ou additives, de type « remplisseurs » (*fillers*) et

« de début ou fin d'énoncés » (*sentence-initial or final,* tels que *et, alors, puis, donc*) d'une part, et d'autre part (2) les conjonctions qui nécessitent un traitement syntaxique particulier pour être placées de manière adéquate à l'intérieur de la chaîne syntagmatique. Il semble en effet que ce que Menn et Obler (1990) appellent « morphèmes non lexicaux optionnels » soient, en fait, des particules de discours qui gravitent en marge de la structuration interne ou « phrastique » de l'énoncé produit. Dans de nombreux cas, des adverbes (tels que *puis, alors, après, voilà,* etc...) de locutions conjonctives ou adverbiales (telles que *et alors, et puis, c'est-à-dire, en fait, par exemple, etc…*) viennent initier, clore ou connecter des énoncés entre eux.

La segmentation ne doit pas corrompre la structuration syntaxique solidaire abstraite, c'est-à-dire, la structure sous-jacente phrastique, même en présence de pauses longues au milieu d'une structure syntaxique.

Aussi, les mots *et* ou *puis* peuvent avoir soit une valeur de connecteur discursif inter-énoncé (au même titre que *alors, en fait*, etc.) et donc il y a lieu de segmenter, soit une valeur de conjonction de coordination intra-phrastique dans la chaîne syntagmatique (« il a mangé et dormi », « il a mangé puis est allé au lit ») et dans ce cas il n'y a pas lieu de segmenter.

Pour la segmentation, la priorité est donc donnée plutôt à la structuration syntaxique et décidée en fonction des marqueurs / connecteurs de discours.

Les problèmes de segmentation sont faciles à résoudre pour les données contrôles. Par contre, pour les aphasies, les phénomènes de pauses longues, d'interruptions fréquentes, d'auto-interruptions, de reformulations, etc. peuvent présenter des difficultés au transcripteur pour segmenter.

En tous les cas, il convient de choisir des critères de segmentation qui resteront rigoureusement les mêmes tout au long du corpus à transcrire.

En cas de doute, il convient aussi de segmenter au plus court !

Voici des exemples de segmentation :

- ***locuteur contrôle - cas ordinaire (les « bullets » en fin d'énoncés sont les repères temporels d'alignement transcription – média avec la fonction F-5) :***

*PAR: le matin je suis allé au parc . ●
*PAR: et puis j'ai profité de l'occasion pour acheter le gâteau que ma femme avait commandé . ●

En cas de chevauchements, terminer la transcription de l'énoncé en question sur la même ligne, même si quelqu'un parle par-dessus, ce qui revient à ignorer le chevauchement. Il existe d'autres codages adaptés pour les chevauchements si on souhaite les coder finement (voir la section « codage »).

- ***locuteur contrôle - 2 énoncés plutôt longs avec des structures complexes :***

*PAR: &-euh j'ai fait un I_R_M pour voir l'état &-euh <plus &+avan> [//] (en)fin de façon plus précise de [//] des dégâts qui avaient été occasionnés par cette chute . ●
*PAR: &-euh et donc très rapidement j'ai entrepris &-euh une rééducation <par le> [//] (.) par un kinésithérapeute puisque le but dans un premier temps c'est de résorber l'œdème et surtout de renforcer musculairement le genou . ●

- ***locuteur aphasique non fluent avec de nombreuses dysfluences et une syntaxe très réduite (les chiffres en fin d'énoncés sont les repères temporels d'alignement à l'audio « bullets étendus » avec la fonction Esc-A ; les chiffres entre parenthèses dans ce corpus correspondent approximativement à des durées de pauses > à 2 secondes) :***

*PAR: &-euh donc &-euh en+fait &-euh (2.) j' ai très mal &au longtemps &euh mal au dos . •0_13443•
*PAR: très longtemps . •13443_14841•
*PAR: et &-hum (2.) &+er <&erdi discale> [//] deux hernies discales . •14841_23247•
*INV: www . •23247_24995•
*PAR: en+fait &-euh &-euh &-euh &-hum (4.) lombaires (2.) et têtes . •24995_35329•
*PAR: les têtes ça xxx résorber . •35329_37947•
*PAR: et &-hum (2.) hernies discales &-hum dos très longtemps &-euh &-euh (2.) coincé &-euh . •37947_51722•
*PAR: donc &-euh (2.) ça va pas+du+tout quoi . •51722_55735•
*PAR: donc &-euh &-euh &-hum (4.) pour ça &-euh &-hum (6.) . •55735_66523•
*PAR: &+te technique &euh trop non &pré présente &euh . •66523_76128•
*PAR: et donc &-euh ça s' appelle professeur Montpellier . •100228_105640•
*PAR: et &-euh bon c' est bon . •105640_109603•
*PAR: et donc &-euh opération &-euh (2.) normalement &-euh (5.) . •109603_116224•
*PAR: &+o opération rien . •116224_121753•
*PAR: <c' est> [/] c' est bon . •121753_122767•
*PAR: sauf [/] sauf &-euh inséstétation@n [//] &-euh anesthésie . •122767_129804•
*PAR: et là réveil . •129804_132899•

2.4.4 Anonymisation des transcriptions et du média

Les informations identifiantes, telles que les noms de personnes ou de lieux trop précis (une adresse par exemple, d'un lieu très précisément donné, etc.) doivent figurer dans la transcription de façon anonymisée en lettres capitales : seules les premières et dernières lettres apparaissent. Une ligne complémentaire %com signale « anonymisation ». Cela permet de retrouver l'occurrence « anonymisation » et l'énoncé concerné dans le fichier audio ou vidéo aligné, et ainsi vérifier si l'audio ou vidéo a bien était remplacé par un bip (vérification de l'anonymisation *a posteriori* possible, facilitée, et plus rapide).

NB : Il n'est pas nécessaire d'anonymiser : si les mots ne sont pas identifiants pour un tiers (les prénoms de personnes, les noms de villes ou de lieux, de région, de lieux publics tels que restaurant, piscine, etc.).

Par exemple : anonymiser les noms de famille ou adresses précises permettant d'identifier précisément des personnes tierses ou le locuteur :

*PAR: le professeur BaylaC dont le cabinet est rue MatabiaU m'a bien soigné .

En modifiant les mots identifiants, et ajoutant la ligne %com : , cela devient :

*PAR: le professeur BC dont le cabinet est rue MU m'a bien soigné .
%com : BC MU = anonymisation

Par exemple : ne pas anonymiser les prénoms, villes, noms d'artistes connus, etc. (ne permettant pas d'identifier le locuteur)

*PAR: j'y ai participé d'ailleurs avec Stéphane un de nos amis .
*PAR: &-euh on est allé à un festival aussi à Lorient avec des amis .
*PAR: et &-euh donc on a vu &-euh bah l'artiste &-euh &=avale <Hubert &+tier> [//] Hubert Felix Thiéfaine .

2.4.5 Codages de base des fichiers .cha

a. Délimiteurs, majuscules, élisions, interjections, « il y a », négation

- Ponctuation : il doit y avoir un délimiteur (signe de ponctuation) à la fin de chaque énoncé. La plupart des énoncés se termine par un point (.), un point d'interrogation (?) ou un point d'exclamation (!). Mettre un espace entre la fin du dernier mot et le signe de ponctuation. Aucun signe de ponctuation ne doit être utilisé à l'intérieur des énoncés (virgules, points-virgules, etc.) à l'exception des apostrophes, exemples : « c'est », « l'enfant ».
- Majuscules : aucune majuscule ne doit être utilisée sauf pour les noms propres (éventuellement avec anonymisation).
- Mots raccourcis ou élisons conformes de l'oralité : parfois, le mot sera raccourci mais toujours intelligible. Mettre la partie supprimée du mot entre parenthèses (phonème, syllabe) : le mot sera ainsi comptabilisé lors des analyses, ce qui est important pour les noms, adverbes, verbes en particulier (mots de classe ouverte) ou les pronoms.

  Exemples :
  (en)chanté (et non « chanté »)
  (en)fin (et non « chanté »)
  j(e) fais à dîner
  t(u) as mangé ?

- Interjections : ces phénomènes typiques et très fréquents à l'oral doivent être transcrits de façon systématique et harmonisée. Les interjections sont morphologiquement étiquetables et reconnues comme mots « communicateurs » sous CLAN.

  Exemples : hein, ouais, nan, ben, beh, eh bien, eh_ben, éh, ah, oh, ouh, ah_bon, pff, badaboum, gla_gla_gla…

  Les interjections sont à distinguer des remplisseurs (fillers) :

  *PAR1: ben demain on fera autre chose hein ?
  *PAR1: ah ouais !

Lors des analyses, il sera possible de mettre à jour les listes de communicateurs, dont les interjections et onomatopées, même s'ils ne sont pas connus par le système :

*PAR: gla-gla-gla &=rit

- Cas de « il y a » : il s'agit d'un figement syntaxique (un présentatif). Dans tous les cas, de façon à harmoniser le codage, transcrire « **il_y a** » :

- qu'il s'agisse d'une prononciation avec élision du « il » = [ja] = « y'a », transcrire « il_y a ». Dans ce cas, il s'agit d'un figement idiomatique correct (usage de l'oralité).
- ou qu'il s'agisse d'une prononciation avec « il » = [ilja], transcrire aussi « il_y a ».

En cas de négation :

- transcrire « il_y a pas » si la négation « n' » n'est pas clairement prononcée ;
- transcrire « il_y en avait déjà plus besoin » si la négation « n' » n'est pas clairement prononcée
- transcrire « il n'y a pas » si la négation « n' » est prononcée clairement ;
- transcrire « il n'y en avait déjà plus besoin » si la négation « n' » est prononcée clairement.

- Cas de la double négation « ne pas », « ne plus », « ne jamais », etc.

A l'oral, la particule « ne » est très fréquemment supprimée, cela est admis par l'usage ordinaire, notamment en conversation. Si elle n'est pas produite, ne pas la transcrire. Si elle l'est, la transcrire, seulement quand elle est clairement prononcée :

*PAR: je pense pas avoir mis en place une gestion particulière de cette problématique &-euh.
*PAR: et &-euh ben je n'ai rien fait de plus .

b. Codages conventionnels .cha et flexibilité

Les transcriptions primaires se conforment aux conventions de segmentation et codages explicitées dans ce manuel. Les corpus de données transcrites respectent les codages de bases décrits dans le tableau synthétique ci-après (les phénomènes d'intérêt, les codages et des exemples).

Cependant, il faut savoir que les analyses *a posteriori* des transcriptions primaires peuvent faire l'objet d'une re-vérification des codages, de corrections et d'un enrichissement éventuel, en fonction des objectifs des analyses envisagées. L'objectif premier de la base de données est de fournir des corpus déjà recueillis et annotés (sur la ligne principale seulement), mais les analystes sont libres de réutiliser et modifier les données ainsi transcrites en fonction d'objectifs ou de questions de recherche spécifiques, et en citant ce manuel en référence.

Toute latitude est ainsi donnée pour les recodages nécessaires des transcriptions, sur ligne principale ou complémentaire, et il est possible de rechercher ou de remplacer un codage par un autre efficacement avec les fonctions de recherche / remplacement textuels.

Phénomène d'intérêt	Codage par convention	Remarques et exemples
Séquence non-traitée	www	Partie de la transcription qui est sur le média, mais que le transcripteur décide de ne pas transcrire (aparté inutile, interruption de l'entretien, etc.) *INV: www . *PAR: www . *INV: très bien poursuivons . NB : les digressions intéressantes mais sans lien avec la tâche donnée peuvent être transcrites, mais ajouter une ligne supplémentaire la signalant : %com: digression .
Mot / séquence inintelligible	xxx	Mots inintelligibles avec une forme phonétique peu claire. Il n'est pas possible de transcrire, on ne peut pas inférer ce qui est dit non plus. *INV: je suis d'accord . *PAR: avec d'autres xxx oui . *INV: très bien poursuivons .
Fin de l'énoncé	. ! ou ?	Délimiteur (signe de ponctuation) à la fin de chaque énoncé : ! ou ? lorsque l'intonation est très marquée. La présence d'un délimiteur est obligatoire. Mettre un espace entre la fin du dernier mot et le signe de ponctuation. *PAR: et ce prince charmant est en [//] à la recherche de l'âme sœur . *PAR: donc il donne un bal . *PAR: et toutes les jeunes filles du royaume sont conviées à ce bal .
Pause vide	(..) (…) (6.)	Marquer les pauses de 2 secondes (..) ou supérieures à 3 secondes (…). Si l'on souhaite être plus précis, marquer la durée (6.) pour 6 secondes par exemple. *PAR: donc &-euh (..) ça va pas+du+tout quoi . *PAR: donc &euh &euh &hum (4.) pour ça &euh &hum (6.) .
Pause remplie	&-	Mot qui ne fait pas partie du lexique conventionnel : une pause remplie (filler) - &-euh &-bah &-ben &-mm &-hum &-eh &- &-beh : *PAR: donc &-euh (..) ça va pas+du+tout quoi . *PAR: donc &euh &euh &hum (4.) pour ça &euh &hum (6.) .
Non-mot, mot déformé néologismes (plus de 50 % du mot déformé ou non reconnaissable)	&	Mot qui est déformé par rapport aux mots du lexique conventionnel &modélité. Si la cible est connue, on peut la préciser à côté entre crochets [: mobilité]. *PAR: ou &-euh &prodalité [: mobilité] très réduit hein &prifsé.
Amorce de mot Fragment phonologique	&+	Amorce de mot ou fragment : *PAR: en fait j'ai [//] <je &+p> [//] je pense que &+j j'ai vite &+c &+comp compris qu'il_y avait une part de stress &-euh qui est [//] pouvait influer .
Répétition simple de mot (sans changement)	[/]	Mot qui est répété une fois sans changement : *PAR: c'est une [/] une fleur [/] fleur .
Répétition de mot multiple (sans changement)	[x N] (même sens que [/] mais avec multiplicateur)	Mot répété plusieurs fois sans changement. *PAR: c'est une [/] une fleur [/] fleur [/] fleur . Équivaut à : *PAR: c'est une [/] une fleur [x 2] .
Répétition de syntagme (plusieurs mots) sans changement	< > [/]	Suite de mots (syntagme) répétée sans changement. La portée de la répétition est marquée par < > . *PAR: <j'ai> [/] j'ai vu une fleur .
Répétition d'un mot avec changement : auto-correction	[//]	Mot effectivement produit et répété avec changement (auto-correction) *PAR: j'ai vu une chose [//] fleur .
Répétition de syntagme (plusieurs mots) avec changement (révision)	< > [//]	Suite de mots (ou syntagme) répétée avec changement. La portée de la répétition est marquée par < > . *PAR: j'ai vu une [/] &-euh <une fleur> [//] &-euh une fleur blanche .
Reformulations	[///]	[///] (à la différence de [/] et [//]) est utilisé pour coder une reformulation complète d'un énoncé sans reprise de ce qui a été dit avant : *PAR: <je voulais> [///] &-euh il était complètement en retard .
Auto-interruption avec changement complet : continuation sur une autre idée	[/-]	[/-] (à la différence de [/], [//] et [///]) est utilisé pour coder un énoncé incomplet qui s'interrompt et se poursuit avec un nouvel énoncé qui change complètement : *PAR: <je voulais> [/-] &-euh quand vient-elle ?
Indications paralinguistiques	&=rire &=geste	Indications de comportements co-verbaux / para-linguistiques. Préciser le phénomène librement : *PAR: le dessin est mal fait &=geste &=regarde-le-dessin . *PAR: oui &=acquiesce *PAR: on ne voit pas la fleur &=rit ! Possibilités : &=rit &=soupire &=avale

		&=déglutit &=claque-la-langue &=regarde-l'image &=siffle
Silence verbal	0 &=silence	Rien n'est dit par le locuteur (0 si rien n'est dit, et éventuellement ce qui se passe) : *PAR: 0 &=silence &=rit .
Élision de phonèmes ou partie d'un mot (élision attestée à l'oral)	(en)fin j(e) , t(u), i(l)	Le mot fait partie du lexique conventionnel, et sa forme à l'oral est habituelle et acceptable car liée à l'usage (réduction de la prononciation sans que ce soit une réelle déformation) : *PAR: (en)fin le dessin est mal fait là ! *PAR: comme j(e) te dis . *PAR: c'est un hélico(ptère) ! *PAR: i(l)_y avait un monde fou (en)fin . *INV: mais t(u) es revenu ?
Mots composés	orthographe ordinaire	Mettre le tiret - quand le mot composé est présent dans le lexique conventionnel à l'écrit : lave-vaisselle est-ce que qu'est-ce que c'est-à-dire
Morphèmes solidaires composés Acronymes composés	parce_que T_G_V A_V_C	Mettre le signe _ pour les noms de personnes ou de lieux, acronymes, morphèmes solidaires / complexes qui n'ont pas de tiret selon les conventions orthographiques, mais qui sont morphologiquement solidaires à l'oral (figements) : *PAR: l'histoire se passe à la La_Pointe_du_Raz *PAR: l'auteur est Simone De_Beauvoir *PAR: il a perdu son bagage dans le T_G_V en partance pour Cesson_Sévigné en Ille_et_Vilaine quand_même ! Transcrire ainsi : parce_que, quand_même Rappel : il_y a
Autres codages possibles présents dans les transcriptions :		
Onomatopées	@o	vroum@o
Interjections	@i	ah@i oh@i hein@i oh_la_la@i pouf@i ouf@i wouah@i han@i pfou@i
Prononciation de lettres Épeler un mot	@l	*PAR : j'ai écrit v@l a@l i@l r@i
Mots composés (autres possibilités de codage)	xxx_zzzz	week_end belle_mère demi_soeur petit_enfant peut_être porte_manteau pique_nique cerf_volant arrière_plan
Chevauchements	<blabla> [>] bla [<]	INV : <tu vas> [>] lui chercher de la colle. PAR : <moi je> [/] [<] moi je ne vais pas en bas
Silence verbal	0	*INV: que pensez-vous de cette image ? *PAR : 0 .
Dyslfuences / phénomènes atypiques	Voir en fonction de la précision du codage envisagée	Concernant les dysfluences, notamment pour les données d'aphasie, voir aussi le chap. 13 *Dysfluency Transcription* du manuel (Mac Whinney, 2021, https://www.talkbank.org/manuals/CHAT.pdf .) Concernant les phénomènes plus précisément identifiés comme atypiques et que l'on souhaite coder / annoter dans la transcription, voir aussi : Bernstein Ratner, N. & Brundage, S.B. & Fromm, D. (2020), A Clinician's Complete Guide to CLAN and PRAAT, updates by D. Fromm : https://talkbank.org/manuals/Clin-CLAN.pdf.

2.4.6 Codages avancés des fichiers .cha

D'autres codages peuvent être ajoutés en fonction des nécessités de travail d'analyses de données. Dans ce cas, il convient de les ajouter aux transcriptions s'ils ne sont pas présents dans la transcription primaire :

1) Chevauchements (en particulier pour la tâche d'interaction) :

- [>] indique un recouvrement avec l'énoncé suivant : figure directement après le mot ou groupe de mots sur lequel porte le chevauchement prononcé par le premier interlocuteur
- [<] indique un recouvrement avec l'énoncé précédent : figure directement après le mot ou groupe de mots sur lequel porte le chevauchement prononcé par le second interlocuteur

  Si le chevauchement porte sur plusieurs mots à la fois ; les mettre entre chevrons < et > :

  *INV: <tu vas> [>] lui chercher de la colle .
  *PAR: <moi je> [/] moi je vais pas en bas [<] .

2) Autres codages possibles :

- les codages utiles en analyse de la conversation (CA pour Conversation Analyses), tels que des marqueurs plus fins de chevauchement, voir sur la page dédiée https://ca.talkbank.org/codes.html (voir aussi le manuel https://www.talkbank.org/manuals/CHAT.pdf, *chap. 12 CHAT-CA Transcription,* Mac Whinney, 2021) ;
- les codages tels que les délimiteurs spéciaux de fin d'énoncés (e. g. auto-interruptions) voir dans le manuel https://www.talkbank.org/manuals/CHAT.pdf, *chap. 14 Transcribing Aphasic Language* (Mac Whinney, 2021), les codages tels que les délimiteurs spéciaux de fin d'énoncés (e. g. auto-interruptions) ;
- les codages tels que les erreurs de type phonologique, syntaxique, morphologique, grammatical, jargon, omissions (0det, 0aux, 0pro, etc.), persévérations, circonlocutions, etc. portant sur le mot ou l'énoncé, voir dans le manuel https://www.talkbank.org/manuals/CHAT.pdf, *chap. 18 Error coding* (Mac Whinney, 2021).

2.4.7 Fonction F5 : alignement entre le fichier .cha et le signal audio-vidéo

Les transcriptions sont alignées au signal audio-vidéo grâce à la fonction F5 https://www.talkbank.org/manuals/CHAT.pdf, *chap. 10, section 1 Audio and vidéo time marks* (Mac Whinney, 2021). Les marqueurs de temps peuvent être étendus avec la touche Esc-A : les temps doivent se suivre pour un alignement correct (sinon le fichier ce sera pas conforme et il faudra corriger les marqueurs de temps pour qu'ils se succèdent).

Exemple :

*PAR: ben en fait &-euh je pense que tout a commencé &-euh par &-euh la poursuite d'un chat . ●
*PAR: &-euh donc un chien et un chat se sont croisés . ●
*PAR: le chien s'est mis a poursuivre le chat . ●
*PAR: le chat affolé a grimpé dans un arbre pour se mettre à l'abri des crocs du chien . ●

Avec les bullets étendus Esc-A :

*PAR: ben en fait &-euh je pense que tout a commencé &-euh par &-euh la poursuite d'un chat . •373755_379580•
*PAR: &-euh donc un chien et un chat se sont croisés . •379580_383317•
*PAR: le chien s'est mis a poursuivre le chat . •383317_384918•
*PAR: le chat affolé a grimpé dans un arbre pour se mettre à l'abri des crocs du chien . •384918_390379•

2.5 Conservation des fichiers

Le serveur Human-Num est dédié au dépôt des fichiers .cha et audio / vidéo, *via* le consortium AADI.

Les fichiers conservés seront réutilisés et modifiés : leur nouvelle forme et de nouveaux fichiers peuvent être déposés en plus dans la base (par ex. un fichier avec la ligne %MOR: générée ou des transcriptions enrichies), dans des dossiers bien distincts.

2.6 Étiquetage %MOR et traitement automatique

Les procédures d'analyses automatiques des fichiers .cha sont décrites sur Talkbank.org (par ex., concernant l'analyse de discours https://aphasia.talkbank.org/discourse/) consulter le manuel CLAN dédié régulièrement mis à jour.

Le système CLAN permet d'étendre les codages avec de nouveaux codages à créer *de novo* dans le système, et à l'aide de lignes complémentaires. L'interopérabilité entre logiciels (CLAN, ELAN, PRAAT, etc.) augmente également les potentialités de codages complémentaires et d'analyses.

3 Références

Aphasiabank—Talkbank (s. d.). https://aphasia.talkbank.org/

Bernstein Ratner, N. & Brundage, S.B. & Fromm, D. (2020), A Clinician's Complete Guide to CLAN and PRAAT, updates by D. Fromm : https://talkbank.org/manuals/Clin-CLAN.pdf.

Colin, C., & Le Meur, C. (2016). *Adaptation du projet AphasiaBank à la langue française – Contribution pour une évaluation informatisée du discours oral de patients aphasiques. Sous la direction de : H. Sahraoui et K. Labrunée-Prod'homme* [Mémoire présenté en vue de l'obtention du Certificat de Capacité d'Orthophoniste, Université Paul Sabatier Toulouse III].
https://aphasia.talkbank.org/access/French/Aphasia/ColinLeMeur.html

Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Lea & Febiger.

Goodglass, H., Kaplan, E., & Barresi, B. (2001a). *Boston diagnostic aphasia examination (3rd ed.)*. Lippincott Williams & Wilkins.

Goodglass, H., Kaplan, E., & Barresi, B. (2001b). *The assessment of aphasia and related disorders (3rd ed.)*. Lippincott Williams & Wilkins.

Hammelrath, C. (1999). *Test de Dénomination de Verbes Lexicaux en images DVL 38*. Ortho Edition.

Hammelrath, C., Rotru, R., & Wilhelm, S. (2000). DVL 38 : Élaboration et standardisation d'un test de dénomination de verbes lexicaux. *Glossa*, *73*, 16-28.

Kerbrat-Orecchioni, C. (2013). *Le discours en interaction*. Armand Colin.

Lalain, M., Pouchoulin, G., Priego-Valverde, B., Pinto, S., (2021). « De la protection des données à la protection de la personne : Réflexions sur l'impact des nouvelles réglementations sur la collecte des corpus », Corpus [En ligne], 22 | 2021, mis en ligne le 15 février 2021, consulté le 16 février 2021.URL : http://journals.openedition.org/corpus/5895 ; DOI : https://doi.org/10.4000/corpus.5895

Lecours, A. R., Nespoulous, J.-L., & Joanette, Y. (1996). *Protocole Montréal-Toulouse MT86*. Ortho Edition.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank : Methods for studying discourse. *Aphasiology*, *25*(11), 1286-1307. https://doi.org/10.1080/02687038.2011.589893

MacWhinney, B. (2021). *The CHILDES Project: Tools for Analyzing Talk*. 3 rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates [on-line updated 2021 version: https://talkbank.org/manuals/CHAT.pdf]

Mazaux, J.-M., & Orgogozo, J.-M. (1982). *Échelle d'évaluation de l'aphasie HDAE (adaptation française du Boston Diagnosis Aphasia Examination, H. Goddglass & A. Kaplan, 1972)*. Editions Scientifiques et Psychologiques.

Menn, L. & Obler, L. K. (1990). Cross-language data and theories of agrammatism, chapter 20. In Menn, L. & Obler, L. K. (Eds.), *Agrammatic aphasia: A cross-language narrative sourcebook* (1369-89). Amsterdam: Benjamins.

ORTOLANG. (s. d.). *Outils et Ressources pour un Traitement Optimisé de la LANGue. Programme « Investissements d'avenir » (ANR–11–EQPX–0032)*, https://www.ortolang.fr/

Ossewaarde, R., Jonkers, R., Jalvingh, F., & Bastiaanse, R. (2020). Quantifying the Uncertainty of Parameters measured in Spontaneous Speech of Speakers with Dementia. *Journal of Speech, Language, and Hearing Research*.

TGIR Human-Num. (s. d.). *Très Grande Infrastructure de Recherche : Humanités Numériques. Https://www.huma-num.fr/* [Archives], https://documentation.huma-num.fr/content/23/211/fr/huma_num-box-presentation-globale.html