FICHE MÉTHODOLOGIQUE OPEN-ACCESS (CC BY 4.0) • SCIENCES HUMAINES & SOCIALES

Conventions de Transcription d'Entretiens Qualitatifs, Pré-Balisage CAQDAS (NVivo, MAXQDA, Atlas.ti, IRaMuTeQ) & Pseudonymisation RGPD

Auteur : J. Terral — Pôle Méthodologie & Révision, Dago Transcription (retranscriptionaudio.com) • **DOI :** 10.5281/zenodo.24224772 • **Licence :** CC BY 4.0

1. Choisir le Niveau de Transcription Adapté à l'Objet de Recherche

En sociologie, science politique, anthropologie, psychologie clinique ou sciences de gestion, le choix de la convention de transcription engage directement la validité empirique du matériau analysé. Trois standards coexistent dans les laboratoires francophones :

Régime de transcription	Règles linguistiques appliquées	Disciplines & usages recommandés	Volume moyen (1 h d'audio)
1. Verbatim intégral strict	Restitution mot à mot exhaustive : hésitations (*euh*), faux départs, répétitions, lapsus, silences minutés, rires, soupirs et chevauchements.	Analyse du discours, psychologie clinique, sociologie des affects, auditions sensibles.	~9 200 à 10 500 mots (~24 à 27 pages A4)
2. In Extenso sociologique épuré	100 % des propos et du lexique de l'enquêté conservés, avec suppression des scories orales involontaires (*euh* répétitifs, bégaiements mécaniques) qui gênent la lecture analytique.	Sociologie des organisations, science politique, sciences de l'éducation, gestion, santé publique.	~8 500 à 9 200 mots (~21 à 23 pages A4)
3. Compte-rendu thématique horodaté	Structuration par séquences thématiques avec citations verbatim exactes entre guillemets sur les segments clés et timecodes réguliers.	Pré-enquête exploratoire, entretiens d'experts cadrés, monographies rapides.	~3 000 à 4 000 mots (~8 à 10 pages A4)

2. Grille Normalisée des Signes de Transcription

L'homogénéité typographique sur l'ensemble des entretiens d'un même corpus est indispensable pour éviter les biais de requêtes textuelles et de comptage lexical :

Phénomène acoustique ou interactionnel	Notation recommandée	Justification technique CAQDAS
Changement de tour de parole	`Enquêteur :` / `Enquêté_07 :` (paragraphe distinct)	Permet le codage automatique par locuteur dans NVivo et MAXQDA sans fusionner deux interventions.
Pause courte vs silence prolongé	`…` (< 2 s) et `(silence 4 s)` (≥ 2 s)	Distingue la respiration syntaxique d'une hésitation réflexive ou d'une retenue émotionnelle.
Manifestations non verbales audibles	`(rires)`, `(soupir)`, `(toux)`, `(ton ironique)`	Placées entre parenthèses en minuscules pour ne pas être confondues avec une coupure éditoriale.
Passage inaudible ou incertain	`[inaudible 00:24:18]` ou `[mot incertain ?]`	Les crochets signalent l'intervention métalinguistique du transcripteur et facilitent la réécoute ciblée.
Interruption / chevauchement	`/` en fin d'énoncé interrompu + `[voix superposées]`	Documente la dynamique interactionnelle (notamment en focus group ou entretien collectif).

3. Règles Techniques de Pré-Formatage par Logiciel (CAQDAS & Lexicométrie)

- **NVivo (Windows / Mac) :** Utiliser rigoureusement la même graphie d'étiquette locuteur du début à la fin du corpus (ex. `Enquêté_07 :` sans variante d'accent ni d'espace) ou appliquer un style Word `Titre 3` dédié aux identifiants de locuteurs afin d'activer la fonction *Auto-Code by Speaker / Style* dès l'importation du `.docx`.
- **MAXQDA & Atlas.ti :** Proscrire les sauts de ligne manuels (`Maj + Entrée`) au profit de vrais paragraphes Word. Insérer des repères temporels normalisés `[HH:MM:SS]` à chaque relance afin de lier directement la bande audio MP3/WAV au texte.
- **IRaMuTeQ & Alceste (analyse textométrique) :** Débuter chaque entretien par une ligne de variables étoilées sans espace interne ni ponctuation : `**** *id_07 *genre_F *age_35_44 *statut_titulaire`. Bannir absolument le caractère astérisque `*` et le tiret cadratin collé dans le reste du corpus.

4. Protocole de Pseudonymisation Conforme RGPD & Comités d'Éthique

Conformément au Règlement (UE) 2016/679 (RGPD) et aux référentiels des Délégués à la Protection des Données (DPO CNRS, INSERM, universités), un entretien audio constitue une donnée à caractère personnel (empreinte vocale + informations biographiques indirectement identifiantes). Trois mesures cumulent sécurité juridique et richesse sociologique :

1. **Pseudonymisation contextuelle raisonnée dès la transcription :** Ne pas remplacer les entités par des `[XXX]` opaques qui détruisent l'intelligibilité sociologique, mais par des équivalents fonctionnels constants définis dans une table de correspondance séparée : `[CHU_Régional_1]`, `[Métropole_Ouest]`, `[Syndicat_Majoritaire]`, `[Collègue_Cadre_A]`.
2. **Cloisonnement de la table de ré-identification :** Conserver le tableau reliant `Enquêté_07` à l'identité civile sur un support chiffré distinct du dossier contenant les fichiers `.docx` partagés au sein de l'équipe de recherche.
3. **Encadrement de la sous-traitance de transcription (Art. 28 RGPD) :** En cas d'externalisation, exiger un accord de confidentialité (NDA) couvrant individuellement chaque audiotypiste et réviseur habilité, interdisant formellement tout réentraînement de modèles d'IA tiers et prévoyant la purge définitive des enregistrements sous 30 jours maximum après livraison.

5. Budgétisation d'un Corpus d'Enquête (Écoles Doctorales, ANR, Chorus Pro)

Sur la base d'un débit francophone moyen de **150 mots/minute** (~9 000 mots/heure), un corpus doctoral de **20 entretiens d'une heure** représente environ **180 000 mots** (~450 pages Word A4) et requiert entre **100 et 130 heures de saisie autonome** s'il est transcrit par le chercheur lui-même (ou 45 à 60 heures de réécoute après pré-transcription IA en 2 passes, contre 0 heure interne en délégation complète à 1 € HT/min avec double relecture humaine croisée).

Pour estimer instantanément et localement (sans téléversement serveur) la durée cumulée exacte de vos fichiers MP3/WAV, le nombre de mots et le budget sur bon de commande universitaire / Chorus Pro :

- **Calculateur gratuit 100 % local :** https://retranscriptionaudio.com/calculateur-transcription-audio/
- **Téléchargement des gabarits Word (.docx) gratuits :** https://retranscriptionaudio.com/modeles-gratuits-pv-cse-constat-nda/

Référence bibliographique recommandée (APA / ISO 690) :
Terral, J. — Dago Transcription (2026). *Conventions de transcription d'entretiens qualitatifs (SHS), compatibilité CAQDAS (NVivo, MAXQDA, Atlas.ti, IRaMuTeQ) et pseudonymisation RGPD* [Note méthodologique en libre accès, CC BY 4.0]. Zenodo. https://doi.org/10.5281/zenodo.24224772 • Modèles Word : retranscriptionaudio.com/modeles-gratuits-pv-cse-constat-nda/

Document placé sous licence libre **Creative Commons CC BY 4.0** — Reproduction et diffusion autorisées dans les écoles doctorales, carnets de recherche et supports pédagogiques avec mention de la source : Dago Transcription (retranscriptionaudio.com).