Ce document a été mis en ligne par l'organisme FormaV®.

Toute reproduction, représentation ou diffusion, même partielle, sans autorisation préalable, est strictement interdite.

Pour consulter tous nos sujets d'annales et corrigés de la Licence LLCER, veuillez visiter :
licencellcer.fr/annales-sujets-et-corriges

LICENCE

LANGUES, LITTÉRATURES ET CIVILISATIONS ÉTRANGÈRES ET RÉGIONALES

ÉVALUATION DE L'UE C4 : EXPLOITATION DE DONNÉES À DES FINS D'ANALYSE

SESSION 2026

Durée : 2 heures

SPÉCIALITÉ / UNITÉ	COEFF
LANGUES, LITTÉRATURES ET CIVILISATIONS ÉTRANGÈRES ET RÉGIONALES	3, INDICATIF ET À CONFIRMER PAR LES MODALITÉS DE CONTRÔLE DES CONNAISSANCES DE L'ÉTABLISSEMENT

Moyens de calculs et matériel autorisés :

L'usage de calculatrice avec mode examen actif est autorisé. L'usage de calculatrice sans mémoire, « type collège », est autorisé.

Aucun document ou matériel autorisé en dehors de ceux fournis avec le sujet et d'une calculatrice autorisée.

Dès que le sujet vous est remis, assurez-vous qu'il est complet.

Ce sujet comporte 8 pages numérotées de 1/8 à 8/8.

Les documents réponses en fin de sujet sont à rendre avec la copie.

DOSSIER SUJET

MISE EN SITUATION : ANALYSE D'UN CORPUS LINGUISTIQUE : PROJET UNIVERSITAIRE DE LINGUISTIQUE DE CORPUS, RÉALISÉ DANS LE CADRE D'UNE FORMATION LLCER

Information sur le cadre de l'évaluation : les éléments fournis ne permettent pas d'identifier une épreuve nationale commune aux licences LLCER sous l'intitulé « C4 : Exploitation de données à des fins d'analyse ». Le cadre national de la licence ne fixe pas, pour toutes les universités, un sujet unique, une durée ni un coefficient pour cette épreuve. Le présent sujet constitue une proposition d'évaluation d'établissement, à adapter aux modalités de contrôle des connaissances et des compétences applicables.

Vous étudiez la fréquence des marques de première personne dans quatre sous-corpus écrits. Les données ont été extraites et codées selon le protocole décrit dans le dossier technique. Votre analyse devra distinguer les résultats observés de leur interprétation et tenir compte des limites liées à la taille et à la composition des échantillons.

Les calculs peuvent être effectués à la main ou à la calculatrice. Pour comparer des sous-corpus de tailles différentes, utilisez le taux pour 1 000 mots : $t = \frac{n}{N} \times 1000$, où n est le nombre d'occurrences codées et N le nombre de mots du sous-corpus. Arrondissez les taux au dixième et les écarts au centième si nécessaire.

PARTIE 1 : Définition de l'objet d'étude et examen des données (5 points)

Variables, protocole de comptage et qualité des données

Appuyez-vous sur les documents DT 1 et DT 2. Les quatre sous-corpus et les résultats de comptage sont fournis dans le tableau du DT 1.

1.1 (2 points) Identifiez la variable de regroupement et la variable mesurée. Précisez leur nature et indiquez l'unité d'observation utilisée dans le tableau.

1.2 (2 points) Expliquez pourquoi les nombres bruts d'occurrences ne suffisent pas à comparer les sous-corpus. Donnez la formule de normalisation à appliquer et définissez chacun de ses termes.

1.3 (1 point) Relevez une limite du protocole de constitution ou de codage des données susceptible d'affecter la comparaison. Justifiez votre réponse à partir des documents fournis.

PARTIE 2 : Traitement quantitatif des données (7 points)

Calcul des fréquences normalisées et comparaison des résultats

Utilisez les effectifs et les tailles de corpus du DT 1. Les taux sont exprimés en occurrences pour 1 000 mots. Présentez les étapes de calcul et indiquez les unités.

Données de comptage fournies

Sous-corpus	Langue	Genre	Taille (mots)	Occurrences de la marque codée
A	Anglais	Littéraire	12 000	240
B	Anglais	Presse	15 000	90
C	Français	Littéraire	10 000	180
D	Français	Presse	14 000	70

Source : jeu de données constitué pour cette évaluation, selon le protocole du DT 2.

2.1 (3 points) Calculez le taux pour 1 000 mots des sous-corpus A, B, C et D. Complétez le tableau du DR 1 et arrondissez chaque résultat au dixième.

2.2 (2 points) Pour chaque langue, calculez l'écart entre le taux du sous-corpus littéraire et celui du sous-corpus de presse. Indiquez le sens de l'écart et son unité.

2.3 (2 points) Calculez, pour chaque langue, le rapport entre le taux littéraire et le taux de presse. Arrondissez au centième. Interprétez ces rapports sans conclure à une différence statistiquement significative.

PARTIE 3 : Interprétation et présentation des résultats (4 points)

Lecture critique des fréquences et choix d'une représentation

Votre interprétation doit s'appuyer sur les calculs de la partie 2 et respecter les limites du corpus. Les comparaisons entre genres peuvent être formulées à l'intérieur de chaque langue. Une comparaison directe des valeurs entre langues exige de rappeler que les formes codées et les corpus ne sont pas nécessairement équivalents.

3.1 (2 points) Rédigez une conclusion de 80 à 120 mots décrivant le contraste observé entre les deux genres pour chacune des langues. Distinguez explicitement le constat quantitatif de son interprétation linguistique.

3.2 (2 points) Proposez une représentation graphique adaptée pour comparer les quatre taux. Précisez le type de graphique, les variables à placer sur chaque axe, l'unité, une légende nécessaire et un titre informatif. Indiquez une précaution de lecture à faire figurer sous le graphique.

PARTIE 4 : Fiabilité, traçabilité et responsabilité de l'analyse (4 points)

Reproductibilité, limites méthodologiques et usage des données

Appuyez-vous sur les documents DT 2 et DT 3. Les réponses doivent proposer des mesures concrètes, applicables à l'analyse présentée.

4.1 (2 points) Proposez deux actions permettant à une autre personne de reproduire le comptage et de vérifier les résultats. Pour chacune, précisez l'information ou la procédure à conserver.

4.2 (2 points) Identifiez deux précautions à prendre avant de diffuser les textes ou les résultats détaillés. Reliez chaque précaution à un risque pertinent, par exemple les droits sur les textes, la présence de données personnelles ou une généralisation abusive à partir d'échantillons limités.

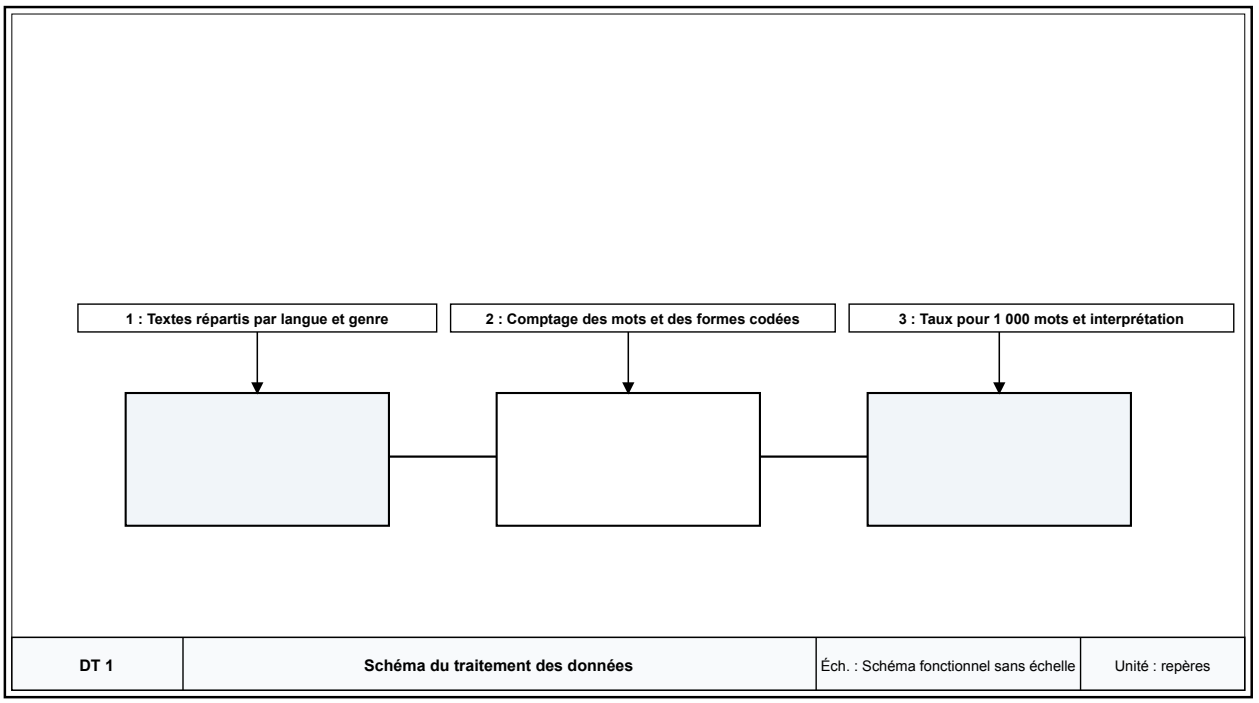
DOSSIER TECHNIQUE ET RESSOURCES

Documents ressources nécessaires à la résolution du sujet

DT 1 : DONNÉES DU CORPUS ET SCHÉMA DE TRAITEMENT

Les nombres de mots et d'occurrences sont ceux indiqués dans le tableau de la partie 2. La marque comptée est, pour chaque langue, une forme de première personne du singulier : dans les textes anglais, le pronom sujet « I » ; dans les textes français, le pronom sujet « je ». Le comptage porte sur les occurrences visibles dans le texte brut, après la préparation décrite au DT 2.

Les taux ont pour formule $t = \frac{n}{N} \times 1000$. Le taux s'exprime en occurrences pour 1 000 mots. Un taux est une mesure descriptive du sous-corpus considéré et ne démontre pas, à lui seul, une relation causale ni une différence statistiquement significative.



Récapitulatif des sous-corpus

Code	Langue	Genre	Nombre de mots	Occurrences
A	Anglais	Littéraire	12 000	240
B	Anglais	Presse	15 000	90
C	Français	Littéraire	10 000	180
D	Français	Presse	14 000	70

La colonne des occurrences correspond à la forme de première personne définie dans l'introduction du DT 1.

DT 2 : PROTOCOLE DE PRÉPARATION ET DE CODAGE

Pour cette évaluation, les textes ont été répartis en quatre sous-corpus selon deux critères : langue et genre. Le nombre de mots est celui communiqué avec chaque sous-corpus. Les occurrences ont été relevées dans les textes préparés en texte brut, sans pondération supplémentaire.

- Anglais : compter chaque occurrence du pronom sujet « I » comme une occurrence de la marque étudiée.
- Français : compter chaque occurrence du pronom sujet « je » comme une occurrence de la marque étudiée. Les formes élidées comme « j'arrive » sont comptées comme une occurrence de « je ».
- Ne pas compter les occurrences citées dans les métadonnées, les titres de fichiers ou les commentaires éditoriaux ajoutés lors de la préparation.
- Le protocole ne distingue pas les fonctions pragmatiques du pronom et ne permet pas, à lui seul, de mesurer la subjectivité, l'engagement ou la position de l'auteur.

Aucun extrait de texte n'est nécessaire pour effectuer les calculs demandés. Les données fournies sont suffisantes pour toutes les questions du sujet.

Règles opérationnelles de comptage

Langue	Forme comptée	Forme élidée	Unité de comptage
Anglais	I sujet	Sans objet	Une occurrence du pronom sujet
Français	je sujet	j' compté comme une occurrence de je	Une occurrence du pronom sujet

Règles appliquées à l'ensemble des sous-corpus pour cette évaluation.

DT 3 : PRINCIPES DE DIFFUSION ET LIMITES D'INTERPRÉTATION

Avant toute diffusion publique de textes ou d'extraits, vérifier les droits applicables aux œuvres et les conditions d'accès aux sources. Si les textes ou leurs métadonnées peuvent contenir des données personnelles, limiter leur diffusion et appliquer les règles de protection des données pertinentes.

Une analyse descriptive doit distinguer les données observées, le protocole de comptage et les interprétations proposées. La composition des échantillons, leur taille, leur période de publication et les différences entre conventions linguistiques peuvent limiter la portée des comparaisons. Un écart de taux observé ne prouve pas, sans analyse complémentaire, que le genre est la cause de cet écart.

Pour permettre la vérification, conserver une version documentée du protocole, les critères de sélection des textes, les étapes de préparation, les paramètres de comptage et les calculs réalisés, dans le respect des droits et des règles de confidentialité.

NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 1 : Document réponse à rendre avec la copie : questions 2.1 et 2.2 **À RENDRE AVEC LA COPIE**

Complétez les cases indiquées par « ... ». Faites apparaître vos calculs sur votre copie et inscrivez les taux arrondis au dixième.

Calcul des taux et des écarts

Sous-corpus	Calcul du taux pour 1 000 mots	Taux pour 1 000 mots
A	.....	.....
B	.....	.....
C	.....	.....
D	.....	.....

Écarts entre genres, calculés au sein de chaque langue

Langue	Taux littéraire	Taux presse	Écart littéraire moins presse
Anglais	.....	.....	.....
Français	.....	.....	.....

Cadre réservé aux justifications, calculs complémentaires ou schémas du candidat :

.....

.....

.....

.....

.....

.....

.....

NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 2 : Document réponse à rendre avec la copie : question 3.2 **À RENDRE AVEC LA COPIE**

Complétez les champs pour décrire la représentation graphique que vous recommandez. Vous n'avez pas à tracer le graphique.

Spécification du graphique

Élément	Réponse du candidat
Type de graphique	.....
Variable de l'axe horizontal	.....
Variable et unité de l'axe vertical	.....
Titre informatif	.....
Légende ou précaution de lecture	.....

Cadre réservé aux justifications, calculs complémentaires ou schémas du candidat :

.....

.....

.....

.....

.....

.....

.....

Copyright © 2027 FormaV®. Tous droits réservés.

FormaV® est une marque déposée auprès de l'INPI sous le numéro 5227897.

Ce document a été élaboré par FormaV® avec le plus grand soin afin d'accompagner chaque apprenant vers la réussite de ses examens. Son contenu (textes, graphiques, méthodologies, tableaux, exercices, concepts, mises en forme) constitue une œuvre protégée par le droit d'auteur.

Toute copie, partage, reproduction, diffusion ou mise à disposition, même partielle, gratuite ou payante, est strictement interdite sans accord préalable et écrit de FormaV®, conformément aux articles L.111-1 et suivants du Code de la propriété intellectuelle. Dans une logique anti-plagiat, FormaV® se réserve le droit de vérifier toute utilisation illicite, y compris sur les plateformes en ligne ou sites tiers.

En utilisant ce document, vous vous engagez à respecter ces règles et à préserver l'intégrité du travail fourni. La consultation de ce document est strictement personnelle.

Merci de respecter le travail accompli afin de permettre la création continue de ressources pédagogiques fiables et accessibles.