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 en savoir plus sur nos formations disponibles, veuillez visiter :
formav.co/explorer

EXPLOITATION DE DONNÉES LINGUISTIQUES

Session 2026 • Code sujet : **26-LIC-SL-EP1** • Durée : **2 heures** • Coefficient : **3**
(indicatif, selon les données fournies)

Sommaire

Cliquez sur une partie pour y accéder directement

1. Examiner le protocole et opérationnaliser la question de recherche (3 questions) **Page 3**

2. Calculer et décrire les fréquences de négation réduite (3 questions) **Page 4**

3. Évaluer l'association entre les variables par un test du khi-deux (4 questions) **Page 6**

4. Analyser des extraits et construire une conclusion critique (3 questions) **Page 8**

5. Ce qu'il faut retenir & Questions fréquentes (FAQ) (Synthèse & Q&A Examen) **Page 10**

Sujet officiel (PDF) : https://licencesl.fr/assets/annales/licence-sl/C2 - Exploiter des données à des fins d'analyse/2026/Licence Sciences du Langage - C2 - Exploiter des données à des fins d'analyse - Session 2026 - 1-Sujet.pdf

1. Examiner le protocole et opérationnaliser la question de recherche

Une équipe de sociolinguistique étudie la variation de la négation en français oral selon le degré de formalité, à partir d'un corpus de 120 occurrences produites par 12 locuteurs. Il faut définir les variables, évaluer les limites du protocole et formuler une hypothèse testable à partir des documents techniques.

Sélection critique des données et définition des variables

Question 1.1. Définissez la variable explicative et la variable étudiée. Pour chacune, indiquez les modalités utilisées dans le dossier et précisez son type de mesure.

La variable explicative est le **contexte d'enregistrement**, avec deux modalités, formel et informel. Elle est qualitative nominale dichotomique. La variable étudiée est la **forme de négation** de chaque occurrence : complète si « ne » ou « n' » est présent, réduite s'il est absent. C'est également une variable qualitative nominale dichotomique. La fréquence de négation réduite est ensuite calculée séparément dans chacun des deux contextes.

> ✓ RÉPONSE À RETENIR
>
> Variable explicative : contexte (formel ou informel). Variable étudiée : forme de négation (complète ou réduite).

Question 1.2. Relevez deux limites du protocole susceptibles de réduire la portée des conclusions et expliquez brièvement l'effet possible de chacune sur l'interprétation.

Première limite : le corpus ne comprend que 12 locuteurs, six par contexte, et les occurrences ont été sélectionnées. Cet échantillon réduit et non représentatif limite la généralisation à l'ensemble des francophones. Deuxième limite : le contexte formel correspond à un récit adressé à un enquêteur, tandis que le contexte informel correspond à une conversation entre proches. La tâche et le type d'interlocuteur varient donc en même temps que la formalité, ce qui empêche d'attribuer avec certitude une différence à la seule formalité.

> ✓ RÉPONSE À RETENIR
>
> Échantillon restreint et sélectionné; formalité confondue avec la tâche et le type d'interlocuteur.

Question 1.3. Formulez une hypothèse testable à partir de la question de recherche et des variables définies.

Hypothèse testable : la fréquence des négations réduites n'est pas la même en contexte formel et en contexte informel. Pour la tester, on compare, dans chaque contexte, le nombre d'occurrences réduites au nombre total d'occurrences de négation. L'hypothèse nulle est que les deux fréquences sont égales; l'hypothèse alternative est qu'elles diffèrent. Cette formulation ne préjuge pas du sens de la différence.

> ✓ RÉPONSE À RETENIR
>
> La fréquence des négations réduites diffère entre les contextes formel et informel.

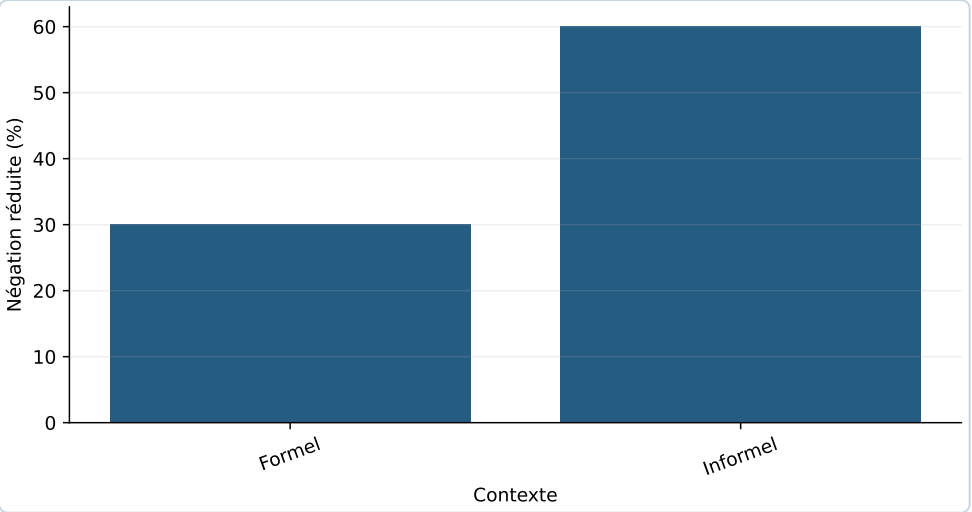
2. Calculer et décrire les fréquences de négation réduite

Le tableau de contingence fournit les effectifs de négation complète et réduite dans deux contextes, formel et informel. Il faut calculer les fréquences de négation réduite, compléter le document réponse et distinguer une différence descriptive d'une conclusion sur l'existence d'un lien statistique.

Calcul de fréquences et comparaison descriptive

Question 2.1. Calculez en pourcentage la fréquence de la négation réduite dans chacun des deux contextes. Présentez les calculs et arrondissez les résultats à une décimale.

La fréquence se calcule en divisant l'effectif de négation réduite par l'effectif total du contexte, puis en multipliant par 100. Pour le contexte formel : $18/60 \times 100 = 30,0\ \%$. Pour le contexte informel : $36/60 \times 100 = 60,0\ \%$. Chaque calcul utilise le total de son contexte, soit 60 occurrences, et non les 120 occurrences du corpus entier.



✓ RÉPONSE À RETENIR

Dans le corpus, la négation réduite représente 30,0 % des occurrences en contexte formel et 60,0 % en contexte informel.

Question 2.2. Complétez le DR 1 pour chaque contexte en indiquant l'effectif de négation réduite, l'effectif total, la fréquence de négation réduite et l'écart entre les deux fréquences en points de pourcentage.

Contexte	Négation réduite	Effectif total	Fréquence	Écart par rapport à l'autre contexte
Formel	18	60	30,0 %	−30,0 points
Informel	36	60	60,0 %	+30,0 points

Les fréquences sont calculées dans chaque contexte. L'écart est signé par rapport à l'autre contexte : le contexte formel est inférieur de 30,0 points au contexte informel, et le contexte informel est supérieur de 30,0 points au contexte formel.

DR 1 complété

Extrait D, situation informelle : « Elle vient pas ce soir, finalement. »

Convention : classer comme négation complète une occurrence où « ne » ou « n' » est présent avec un marqueur négatif; classer comme négation réduite une occurrence où « ne » ou « n' » est absent et où le marqueur négatif est exprimé. Pour cette tâche, analyser uniquement la négation cible selon cette convention.

18	30,0 %	−30,0 points
36	60,0 %	+30,0 points

Convention de codage des exemples

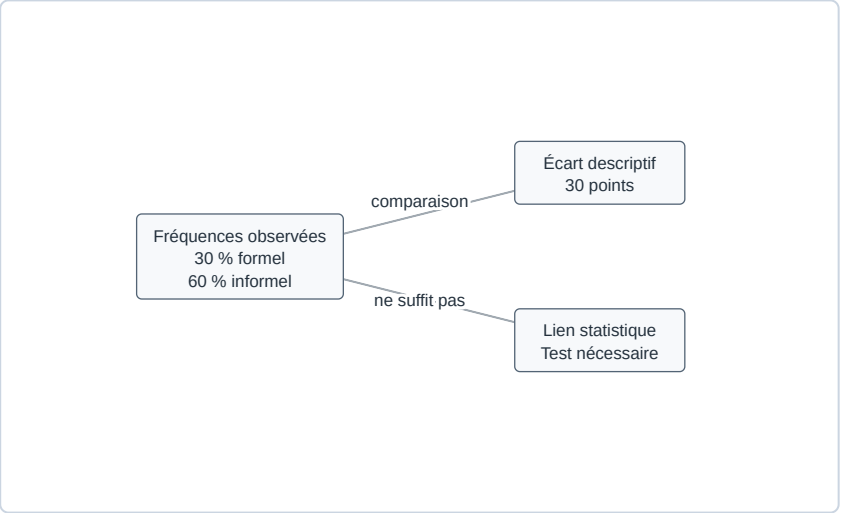
✓ RÉPONSE À RETENIR

Formel : 18 sur 60, soit 30,0 %. Informel : 36 sur 60, soit 60,0 %. Écart entre les fréquences : 30,0 points.

Question 2.3. Décrivez la tendance observée en comparant les deux contextes. Distinguez explicitement la différence descriptive d'une conclusion sur l'existence d'un lien statistique.

La négation réduite apparaît plus souvent dans les occurrences du contexte informel : 36 sur 60, contre 18 sur 60 dans le contexte formel. L'écart descriptif entre les fréquences du corpus est de 30 points de pourcentage. Ces proportions décrivent les données observées, mais ne suffisent pas à établir un lien statistique entre formalité et forme de négation. Cette conclusion nécessite l'examen du test statistique prévu dans la suite du sujet.

Comparer les fréquences observées



Les proportions décrivent le corpus ; elles ne suffisent pas à établir une association statistique.

✓ RÉPONSE À RETENIR

Dans le corpus, la négation réduite est plus fréquente en contexte informel, avec un écart descriptif de 30 points. La conclusion statistique dépend du test.

3. Évaluer l’association entre les variables par un test du khi-deux

Les effectifs observés sont organisés dans un tableau croisant le contexte et la forme de négation. Il faut calculer les effectifs théoriques et la statistique du khi-deux, puis prendre et interpréter une décision au seuil de 5 % à partir de la valeur critique fournie.

Tableau de contingence et test du khi-deux d’indépendance

Question 3.1. Calculez les quatre effectifs théoriques sous l’hypothèse d’indépendance, montrez au moins un calcul détaillé et inscrivez les quatre résultats dans le DR 2.

Sous l’hypothèse d’indépendance, l’effectif théorique vaut : total de ligne × total de colonne ÷ effectif global. Exemple détaillé : pour Formel et négation complète, $60 \times 66/120 = 33$. Les deux lignes ayant le même total, les effectifs théoriques sont identiques pour chaque forme dans les deux contextes. Voici les quatre valeurs à reporter dans le DR 2 :

Contexte	Négation complète	Négation réduite
Formel	33	27
Informel	33	27

DR 2 complété : effectifs théoriques

votre copie.

Effectifs théoriques sous l’hypothèse d’indépendance

Contexte	Négation complète : effectif théorique	Négation réduite : effectif théorique

✓ RÉPONSE À RETENIR

DR 2 : Formel, 33 et 27 ; informel, 33 et 27.

Question 3.2. Calculez la statistique χ² à partir des quatre cellules. Présentez les contributions de chaque cellule et donnez le résultat arrondi à trois décimales.

Chaque contribution se calcule par $(O - E)^2/E$. Les écarts observé-théorique valent 9 dans les quatre cellules. Les contributions sont donc : Formel, complète $81/33 = 2{,}455$; Formel, réduite $81/27 = 3$; Informel, complète $81/33 = 2{,}455$; Informel, réduite $81/27 = 3$. La statistique du khi-deux est la somme de ces quatre contributions, calculée avant l’arrondi final.

✓ RÉPONSE À RETENIR

$\chi^2 = 10{,}909$.

Question 3.3. **Indiquez le nombre de degrés de liberté, comparez la statistique calculée à la valeur critique fournie et formulez la décision au seuil de 5 %.**

Le tableau comporte deux lignes et deux colonnes. Le nombre de degrés de liberté est donc $(2-1)(2-1)=1$. La statistique calculée, $10{,}909$, est supérieure à la valeur critique fournie, $3{,}841$. Selon la règle de décision indiquée dans le sujet, on rejette l'hypothèse d'indépendance au seuil de 5 %. Cette décision concerne le test d'association entre le contexte et la forme de négation.

> ✓ **RÉPONSE À RETENIR**
>
> 1 degré de liberté ; $10{,}909 > 3{,}841$: rejet de l'hypothèse d'indépendance au seuil de 5 %.

Question 3.4. **Interprétez la décision dans le cadre de l'étude. Précisez ce que le test permet de conclure et ce qu'il ne permet pas d'affirmer.**

Dans le corpus, le contexte et la forme de négation sont associés : la négation réduite compte 36 occurrences en contexte informel contre 18 en contexte formel. Le test indique que les données sont incompatibles avec l'indépendance au seuil de 5 %. Il ne démontre pas que la formalité cause cette différence et ne garantit pas sa généralisation à tous les francophones. Le corpus comprend seulement 12 locuteurs, répartis en deux groupes distincts, et les occurrences d'un même locuteur peuvent être liées.

> ✓ **RÉPONSE À RETENIR**
>
> Association observée dans ce corpus, sans preuve de causalité ni généralisation assurée à l'ensemble des francophones.

4. Analyser des extraits et construire une conclusion critique

Quatre extraits anonymisés illustrent la convention de codage des négations complètes et réduites, sans constituer des observations supplémentaires du corpus quantitatif. Il faut les classer, expliquer l'apport limité de ces exemples qualitatifs et rédiger une conclusion argumentée en tenant compte des réserves méthodologiques.

Analyse qualitative, synthèse et argumentation

Question 4.1. Pour chacun des quatre extraits du document DT 3, indiquez si la négation est complète ou réduite. Justifiez chaque classement par un élément linguistique observable et complétez le DR 2.

La convention du DT 3 classe comme complète une négation où « ne » ou « n' » accompagne le marqueur négatif, et comme réduite une négation où ce marqueur est absent. Pour chaque extrait, l'indice cité est directement observable dans l'énoncé. Seule la négation ciblée par la convention est classée. Ces exemples illustrent le codage et ne s'ajoutent pas aux effectifs du corpus.

Extrait	Classement	Indice linguistique
A	Négation complète	« ne pense pas »
B	Négation réduite	« sais pas », sans « ne »
C	Négation complète	« n'a pas »
D	Négation réduite	« vient pas », sans « ne »

DR 2 complété : classement des extraits



✓ RÉPONSE À RETENIR

A et C : négations complètes. B et D : négations réduites, selon la convention du DT 3.

Question 4.2. Expliquez en quoi les extraits qualitatifs peuvent compléter les résultats quantitatifs, sans leur attribuer une représentativité qu'ils n'ont pas.

Les extraits rendent concrète la règle de codage : « Je sais pas » montre une négation réduite, tandis que « Je ne pense pas » montre une négation complète. Ils aident ainsi à comprendre quelles formes sont comptées dans les résultats quantitatifs. Mais le DT 3 précise qu'il s'agit d'exemples construits, et non d'observations supplémentaires du corpus. Ils ne permettent donc ni d'estimer des fréquences ni de représenter les usages des locuteurs. Leur apport est illustratif et méthodologique, pas statistique.

✓ RÉPONSE À RETENIR

Les extraits illustrent la règle de codage, mais, parce qu'ils sont construits, ils n'établissent aucune fréquence représentative.

Question 4.3. Rédigez une conclusion argumentée de 100 à 150 mots répondant à la question de recherche. Mobilisez au moins deux résultats chiffrés ou observations du dossier et formulez une réserve méthodologique pertinente.

Dans le corpus, la négation réduite atteint 60 % en contexte informel (36/60) contre 30 % en contexte formel (18/60), soit un écart descriptif de 30 points. Le test du khi-deux calculé sur le tableau donne $\chi^2(1) = 10{,}909$, supérieur à la valeur critique de 3,841. Selon ce test, les variables sont associées au seuil de 5 %. Cette conclusion concerne les occurrences analysées et n'établit pas un effet causal de la formalité. Les exemples du DT 3, comme « Je sais pas » en contexte formel et « On n'a pas » en contexte informel, illustrent des formes différentes, mais sont construits et ne s'ajoutent pas au corpus. Enfin, les 120 occurrences proviennent de seulement 12 locuteurs, avec 10 occurrences par personne et des groupes distincts. La dépendance possible entre réponses d'un même locuteur fragilise l'hypothèse d'indépendance du test. Ces résultats suggèrent donc une variation selon le contexte dans cet échantillon, sans autoriser une généralisation large.

✓ RÉPONSE À RETENIR

La négation réduite est plus fréquente dans l'informel; l'association observée est significative, mais sa généralisation reste limitée.

SYNTHÈSE DE L'ÉPREUVE & RÉFÉRENTIEL • LICENCE SCIENCES DU LANGAGE

Ce qu'il faut retenir de ce sujet (Session 2026)

Bilan de l'épreuve : Évaluation du bloc RNCP38692BC02, session 2026, code 26-LIC-SL-EP1, d'une durée de 2 heures et de coefficient indicatif 3. L'épreuve étudie la variation de la négation réduite en français oral selon la formalité. Elle évalue l'analyse critique d'un protocole, le calcul de fréquences et d'un khi-deux, le codage d'extraits et la formulation d'une conclusion prudente.

Notions-clés, formules & règles à maîtriser

- Fréquence en pourcentage : effectif de la modalité étudiée divisé par l'effectif total du groupe, puis multiplié par 100.
- Sous l'hypothèse d'indépendance, l'effectif théorique d'une cellule vaut le produit du total de sa ligne et du total de sa colonne, divisé par l'effectif global.
- La statistique du khi-deux est la somme, sur les cellules, de (effectif observé - effectif théorique)² divisé par l'effectif théorique.
- Une négation complète comporte « ne » ou « n' » avec un marqueur négatif; une négation réduite omet ce marqueur tout en conservant le marqueur négatif.

Points de vigilance & attentes du jury

- Présentez les étapes de calcul et respectez les arrondis demandés : une décimale pour les fréquences et trois pour χ^2.
- Exprimez l'écart entre fréquences en points de pourcentage, et non en pourcentage relatif.
- Distinguez la tendance descriptive observée dans le corpus d'une généralisation à l'ensemble des locuteurs.
- Justifiez chaque codage par la présence ou l'absence de « ne » ou « n' » et n'ajoutez pas les exemples construits aux effectifs.

Questions fréquentes sur cette épreuve (Q&A Examen)

Q : Comment calculer et comparer les fréquences de négation réduite dans ce sujet ?

Divisez l'effectif de négation réduite par l'effectif total du contexte, puis multipliez par 100. Les résultats sont 18/60, soit 30 %, en contexte formel et 36/60, soit 60 %, en contexte informel. L'écart est de 30 points.

Q : Comment décider si le test du khi-deux montre une association ?

Calculez les effectifs théoriques à partir des marges, puis additionnez les contributions des quatre cellules. Avec un degré de liberté, comparez χ^2 à 3,841 : s'il la dépasse, rejetez l'hypothèse d'indépendance au seuil de 5 %.

Q : Comment organiser son temps pour cette épreuve de deux heures ?

Répartissez votre temps entre les quatre parties en tenant compte de leurs barèmes : 4, 5, 6 et 5 points. Gardez du temps pour vérifier les calculs, compléter les DR 1 et DR 2 et rédiger la conclusion de 100 à 150 mots.

Q : Quelle réserve méthodologique faut-il considérer dans l'interprétation du test ?

Les 120 occurrences proviennent de seulement 12 locuteurs, avec dix occurrences par personne. Les occurrences d'un même locuteur peuvent être liées, ce qui fragilise l'hypothèse d'indépendance; la petite taille et la sélection du corpus limitent aussi la généralisation.

Sujet officiel associé (PDF) : https://licencesl.fr/assets/annales/licence-sl/C2 - Exploiter des données à des fins d'analyse/2026/Licence Sciences du Langage - C2 - Exploiter des données à des fins d'analyse - Session 2026 - 1-Sujet.pdf
Ressources Licence Sciences du Langage : Toutes les annales & corrigés • Fiches de révision complètes • Simulateur de notes d'examen

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.