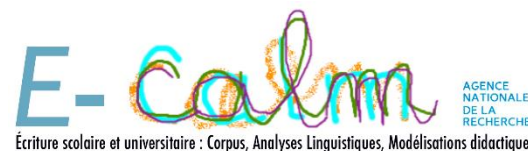


pendant l'uni, E
nt ena réciproc
dant les cours
e, chacune des y
t de une par
téléphone postal
que de red
pour le nom
favaine

Outils de traitement automatique des corpus d'apprenants : méthodes et exploitation

avant acceter
8 <feature nam
9 <possibleV
10 <value>p
11 <value>p
12 <value>p
13 <value>e
14 </possible
15 </feature>
16 </features
17 </type>
18 <type name
19 <featureSet
20 <feature nam
21 <possible
22 <value>
23 <value>
24 <value>
25 <value>
26 </possibl
27 </feature>

Luca Pallanti – Séminaire DeLiCorTAL
17 septembre 2021



Objectifs du séminaire

Le séminaire fournira des pistes méthodologiques pour **l'étude de la cohérence et de la cohésion dans les textes d'étudiants** issus des filières technologiques (L1) à travers, notamment, l'explicitation du processus de conception d'un **modèle d'annotation des chaînes de référence**. De même, nous montrerons une procédure d'annotation visant l'analyse des performances orthographiques. Enfin, l'intervention tentera de comprendre dans quelle mesure la linguistique de corpus outillée peut jouer un rôle dans l'évaluation des dispositifs didactiques expérimentaux axés sur le développement des compétences rédactionnelles.

Plan de la présentation

1. Introduction : le contexte de recherche
2. L'arrière-plan épistémologique
3. Le corpus ÉNEPS : procédures d'annotation
4. Méthodes d'analyse et résultats
5. Conclusion et perspectives
6. Bibliographie

1. Introduction : le contexte de recherche

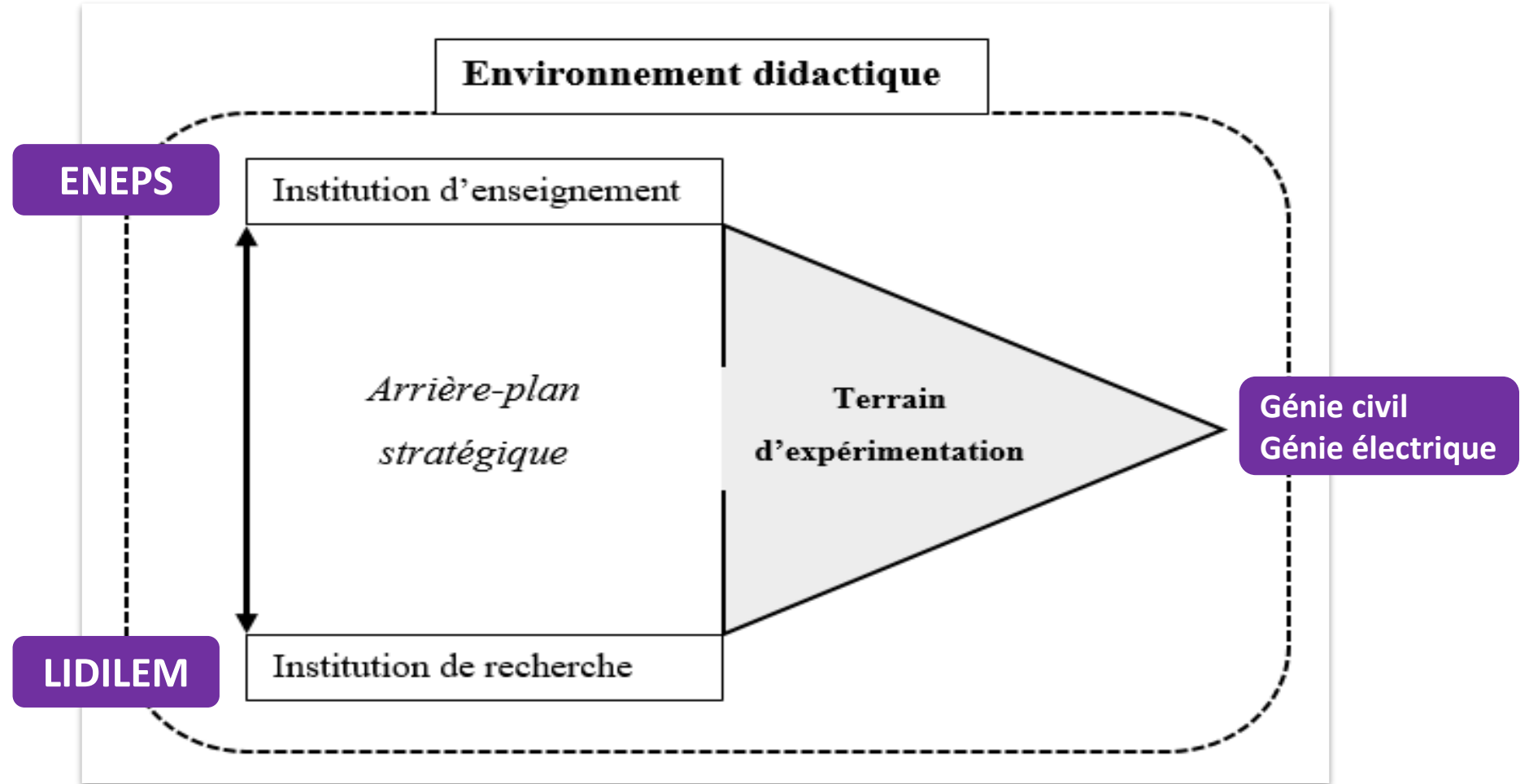
1.1 Objectifs de la thèse

Enjeu : élaborer un cadre de recherche sur les compétences rédactionnelles des étudiants et les écrits en situation de travail pour développer un dispositif didactique adapté au public de l'ÉNEPS

Objectifs :

- Améliorer les compétences rédactionnelles des étudiants de l'ÉNEPS (Pallanti, Jacques & Brissaud, 2021)
- Concevoir et expérimenter un dispositif didactique de l'écriture (Pallanti, Brissaud & Jacques, 2019)
- Mesurer les effets du dispositif à travers un système de pré et post-tests

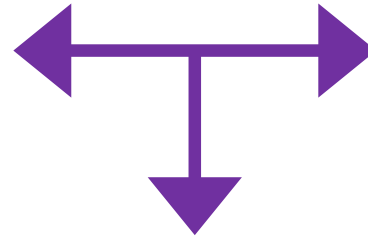
1.2 Modélisation de l'environnement didactique



2. L'arrière-plan épistémologique

2.1 Les champs disciplinaires mobilisés

Champ de la didactique
de l'écrit
(*educational design*)



Champ de la linguistique
appliquée
(*applied linguistics*)

Domaine des littéracies universitaires
(*academic literacies*)

- ❖ Rapport à l'écriture comme fait social et affectif
- ❖ Apprentissage de l'écriture comme fait linguistique et cognitif

2.2 Linguistique appliquée et éducation : quels liens ?

Origine de l'expression : allemand *angewandte Sprachwissenschaft*, et a été employée pour la première fois par Wüster (1931) ➔ *Linguistique appliquée au domaine des sciences techniques*



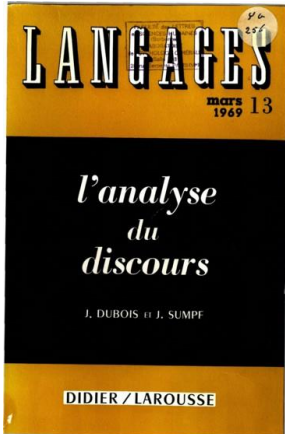
« Nos écoles sont dirigées par des personnes qui, des professeurs d'éducation jusqu'aux maîtres de classe, ne connaissent rien des résultats de la science linguistique, pas même la relation entre l'écriture et la parole, ou entre la langue standard et le dialecte. En bref, ils ne savent pas ce qu'est la langue, et pourtant ils doivent l'enseigner, et par conséquent ils gâchent des années de vie de chaque enfant et obtiennent des résultats médiocres. »

(Bloomfield, 1925, p. 5)

1946

Création de la revue *Language Learning : A Journal of Applied Linguistics*, à l'University of Michigan. Dans la suite des travaux de L. Bloomfield, la linguistique développe une dimension applicative dans le domaine de l'éducation.

2.3 Le champ de la linguistique textuelle



Dans l'article intitulé « Analyse du discours », Harris (1969) fait connaître au public francophone ses théories sur la discursivité, tout en développant parallèlement une linguistique de la textualité. En introduisant une vision distributionnelle des phrases, il contribue de manière décisive à la création du champ de la linguistique textuelle.

Linguistique
textuelle

La linguistique textuelle se donne pour mission l'exploration et la description de phénomènes linguistiques qui comprennent, entre autres, la connexité, la cohésion et la cohérence, qui font qu'une succession d'énoncés peut être considérée comme un texte.

(Bronckart, 2000)

2.4 Évaluer la compétence textuelle

Qu'entend-on par compétence textuelle ?

Charolles : « n'importe quel assemblage de mots ne constitue pas une phrase ». (1978, p.7) **et** « tout tas de phrases ne forme pas un texte » (1978, p.8)

Dans notre étude

Texte en tant qu'espace sémiotique multidimensionnel évalué sous l'angle de la cohérence textuelle

Linéarité textuelle : ordre d'apparition des segments composant le texte

Niveau local (plan *séquentiel*) et **niveau global** (plan *textuel*)

Structure logico-sémantique et plan de la **cohésion textuelle** (relations de *connexité*)

Approche théorique à partir de 4 méta-règles de cohérence (Charolles, 1978)

- Méta-règles de répétition
- Méta-règles de progression
- Méta-règles de non-contradiction
- Méta-règles de relation

2.5 Compétence textuelle et coréférence

Qu'entend-on par référent ?

« la seule condition requise pour que l'on puisse dire d'une expression qu'elle réfère, c'est que son usage dans tel ou tel contexte permette de configurer un point de repère existentiel ancrant une représentation d'une entité du monde externe dans l'esprit des participants à la communication »

(Charolles, 2002, p.33)

« Comme on parle, métaphoriquement, de « chaîne », chacune de ces expressions référentielles constitue un « maillon » de la chaîne. Selon l'approche choisie, certains auteurs utilisent le terme de *chaîne de référence*, d'autres celui de *chaîne de coréférences* (avec également des variations quant au nombre de « références » ou de « coréférences »).

(Schnecker, Glikman & Landragin, 2017, p.6)

**Objectif
de l'étude**

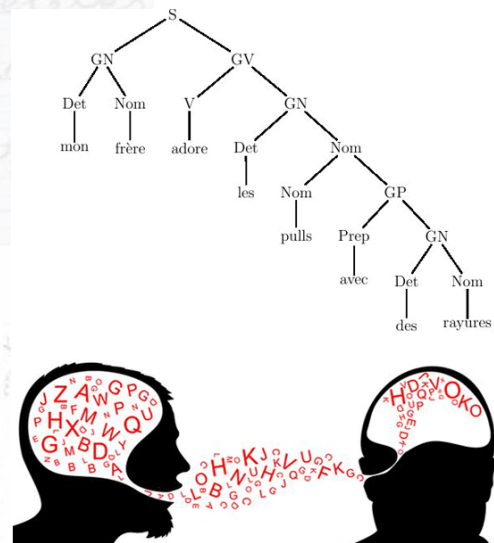
Annotation des chaînes de référence dans un corpus de textes d'apprenants comme clé pour mener une étude de la compétence textuelle

2.6 Les méthodes de la linguistique de corpus outillée

Qu'elle donne des objectifs descriptifs ou applicatifs, la linguistique de corpus des années 2000 se configure comme une linguistique outillée, où l'outillage fait référence à « [...] la mise en œuvre de logiciels pour le traitement des corpus et l'accès de plus en plus facile à des corpus informatisés de plus en plus variés et volumineux ».

(Jacques, 2016, p. 88)

2.7 L'apport du TAL



« Cette méthodologie a pour objet de **décrire comment appréhender les données textuelles** regroupées dans le corpus. Elle inclut la recherche de mots, de portions de mots, d'expressions, avec des possibilités beaucoup plus nombreuses que celles offertes par un logiciel de traitement des textes. Elle intègre également la **méthodologie de l'annotation** – on explore un corpus en l'annotant – et celles de la **linguistique de corpus et des statistiques textuelles**, qui fournissent des indicateurs (numériques, graphiques, multidimensionnels) pour aider l'analyste, linguiste ou non, à mieux comprendre les données constituant son corpus de travail, à mieux caractériser celui-ci, à mieux en dégager les spécificités. »

(Poudat & Landragin, 2017, p. 9)

2.8 Linguistique et corpus d'apprenants



La nature du corpus à constituer et les finalités de la recherche - **mieux comprendre le fonctionnement des textes d'élèves** produits dans un contexte donné - nous a conduites à choisir entre deux conceptions du corpus que F. Rastier (2002) oppose : une **conception logico-grammaticale** comme échantillon de la langue, réservoir d'exemples et d'attestations, et une **conception philologique-herméneutique** où le corpus permet la mise en relation de textes complets, rapportés à un genre et à un discours. Notre hypothèse étant que **les textes d'apprenants sont des textes à part entière**, mais en devenir, d'un état initial à un état jugé final et fortement dépendants d'autres textes, lus, imités, mobilisés dans la séquence d'écriture, le choix de la seconde conception s'imposait.

(Elalouf & Boré, 2007, p. 54)

2.9 Les axes de travail

**Annotation
de la référence
en corpus d'apprenants**

**Etude outillée
des chaînes
de référence
au sein d'un dispositif
expérimental**

**Problématique
des textes éloignés de
la norme et évaluation
des compétences
textuelles**

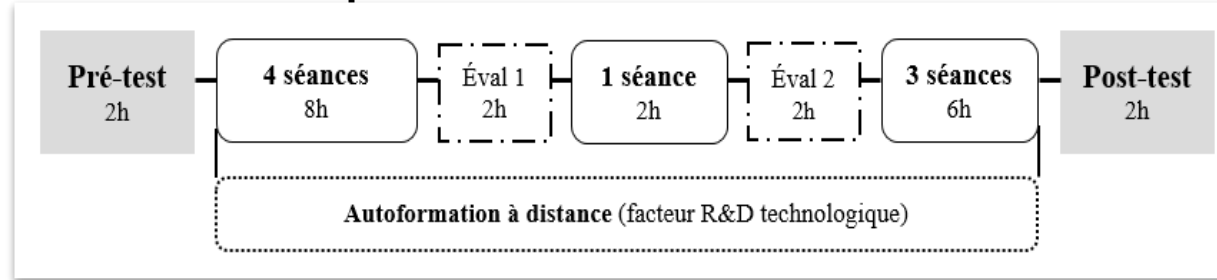
**Linguistique appliquée
et linguistique outillée**

**Didactique
des compétences
rédactionnelles**

3. Le corpus ÉNEPS : caractéristiques et annotation

3.1 Les caractéristiques du Corpus ÉNEPS

Modélisation de l'expérimentation didactique



Consigne du test

Lors d'une enquête réalisée auprès de la population, Alain expose son point de vue. Dans un texte rédigé, vous indiquerez en premier lieu quelle position défend Alain, puis quels arguments il utilise pour ce faire. [...] **Attention : il ne s'agit pas de donner votre avis sur la question.** [...] Votre texte sera évalué sur [...]:

- Indication de la question faisant débat et du positionnement d'Alain à ce sujet

Rédiger une synthèse écrite à partir de l'écoute d'un fichier sonore

donc

- Pré-tests : 2 référents explicitement requis : **Alain** et **économie de partage**
- Post-tests : 2 référents explicitement requis : **Alain** et **manger de la viande**

Rappel du corpus

	Nbre textes	Nbre mots
Pré-test	49	9.700 ≅
Post-test	49	11.800 ≅
Total	98	21.500 ≅

3.2 Conception du modèle d'annotation

Référents
à annoter

Alain

Nom propre

économie de partage

Expression nominale

manger de la viande

Expression prédicative

Caractéristiques
des unités

Maillons formels
simples à identifier:
Pronoms, possessifs...

Sémantique:
Équivalents et reformulations

Exemples à
interpréter

« **Il** n'est plus sûr
qu'il y est des
limitations a cette
souffrance. »
(POST-GC-3)

« Alain dit aussi qu'il faut faire
attention a cette économie, qui a
lieu essentiellement sur internet
[...]. **Ce mode de consommation**
leur permet ainsi d'éviter les
charges. »
(PRE-GE-12)

« [...] il ne faut pas se
priver de viande »
(POST-GC-10)

3.3 Le modèle d'annotation

Référent	Traits	Description
Alain	littéral	Le référent est répété à l'identique ou presque dans le texte.
	équivalent	Le référent est représenté par une expression équivalente.
	pronom	Le référent est représenté à travers l'emploi d'un pronom personnel, relatif ou réfléchi.
	possessif	Le référent est représenté par un déterminant possessif
	ellipse	La présence du référent est évidente malgré l'omission d'éléments linguistiques s'y référant directement.
économie de partage	littéral	Le référent est répété à l'identique ou presque dans le texte.
	équivalent	Le référent est représenté par une expression équivalente.
	réduction	Une partie du référent est reprise de manière littérale ou équivalente.
	pronom	Le référent est représenté à travers l'emploi d'un pronom personnel, relatif ou réfléchi.
	possessif	Le référent est représenté par un déterminant possessif.
	démonstratif	Le référent est remplacé sans ambiguïté par un pronom démonstratif.
	ellipse	La présence du référent est évidente malgré l'omission d'éléments linguistiques s'y référant directement.
manger de la viande	littéral	Le référent est répété à l'identique ou presque dans le texte.
	équivalent	Le référent est représenté par une expression équivalente.
	réduction	Une partie du référent est reprise de manière littérale ou équivalente.
	pronom	Le référent est représenté à travers l'emploi d'un pronom personnel, relatif ou réfléchi.
	possessif	Le référent est représenté par un déterminant possessif.
	démonstratif	Le référent est remplacé sans ambiguïté par un pronom démonstratif.
	ellipse	La présence du référent est évidente malgré l'omission d'éléments linguistiques s'y référant directement.
	quantiPlus	Le référent est caractérisé positivement par une forme de restriction.
	quantiMoins	Le référent est caractérisé négativement par une forme de restriction.
	quantiNeutre	Trait par défaut.

3.4 La préannotation

Exemple de préannotation

```
<?xml version="1.0" encoding="UTF-8"?>
<texte>
- <p>
- <rat>
  <gap/>
</rat>
  Dans cette enquête nous retrouvons
  <ref ident="Alain">Alain</ref>
  qui nous parle de la problématique suivante faut-il encourager l'
  <ref ident="economiePartage">économie de partage</ref>
  ? Pour
  <ref ident="Alain">Alain</ref>
  la réponse est
- <rat>
  <gap/>
</rat>
  oui mais dans certaines limite car de sont point de vue au dela d'un certain seuil de
  revenue il faudrait interdire.
</p>
- <p>
  Pour cela il va avancer plusieurs argument pour
- <rat>
  <gap/>
</rat>
  et contre pour appuier c'est propos. en premier et le plus important il nous dit que
  cela permet de faire des economie car on pa
- <rat>
  <gap/>
</rat>
  <ajout>ye</ajout>
  aune personne qui n' pas pour but le bénéfici
- <rat>
  <gap/>
</rat>
  <ajout>c</ajout>
  e. bien que des application en profite. De plus il avance le fait de la concurrence
  <ajout>deloyale</ajout>
  au entreprise qui doivent repeter les norme et payer les charge.
```

3.5 L'annotation avec Glozz

The screenshot shows the Glozz annotation tool interface. On the left, a text document is displayed with several phrases highlighted in green boxes: "manger de la viande", "se passer de viande", "la viande", "mangeaient de la viande", "consommation de viande", and "de la viande". The phrase "manger moins de viande" is also highlighted with a red dashed box. A red box above the text contains the following information: ID=20, Type=mangerViande, Author=lpallanti, LastModifiedBy=lpallanti. On the right, a table titled "Features" lists various features and their counts. The "litteral" feature has a count of 1, and the "quantiMoins" feature has a count of 1. Below the table, the "Annotation model" section shows a list of units: Alain, economiePartage, mangerViande, incertain, Rature, and Ajout. The "mangerViande" unit is circled in red. Red arrows point from the circled "1" in the "litteral" row, the circled "1" in the "quantiMoins" row, and the circled "mangerViande" unit to the red dashed box around "manger moins de viande" in the text.

Feature name	Featu
litteral	1
equivalent	0
reduction	0
pronom	0
possessif	0
demonstratif	0
ellipse	0
quantiPlus	0
quantiMoins	1
quantiNeutre	0

Annotation model:

Units	Relations	Sche
Alain		
economiePartage		
mangerViande		
incertain		
Rature		
Ajout		

3.6 Les exemples de désaccord

POST-GC-8

Dans ce reportage, Alain nous expose son point de vue sur la question de l'alimentation. Effectivement, faut-il encore manger de la viande de nos jours? Alain aborde trois points autour de ce sujet, d'abord celui de la santé, ainsi que celui de l'écologie et termine avec la souffrance animale.


Alain explique que pour les particuliers c'est un intérêt évident de faire de l'éco-partage car cela permet de dépenser moins.

PRE-GC-7

PRE-GC-12

Alain nous présente son avis sur l'Economie de partage. Pour Alain, Cela est avantageux pour les particuliers. ils pourront faire des économies sur le logement comme B&B. faire du covoiturage, l'une des aspect interessant

3.7 Le guide d'annotation



GUIDE D'ANNOTATION DU CORPUS ÉNEPS

Jun 2020 – Version 7.1

PARTIE 1 : Pour une étude des chaînes des références
PARTIE 2 : Pour une étude de l'orthographe appliquée

Pallanti, L., Jacques, M.-P., Brissaud, C., Guide d'annotation du Corpus ÉNEPS. Version 7.1, Projet E-CALM, Laboratoire LILILEM, UGA, Grenoble Alpes, juin 2020.

Table des matières

- 1 Présentation générale et objectifs du manuel d'annotation 2
 - 1.1 Composition du corpus ÉNEPS 2
 - 1.2 Transcription et pré-annotation 3
 - 1.3 Les outils 3
- 2 PARTIE I : Pour une étude des chaînes de référence 4
 - 2.1 Conventions et définitions 4
 - 2.2 L'annotation des chaînes des références dans le corpus PADE-PRO 4
 - 2.3 Les choix d'annotation : le système des unités 5
 - 2.4 Les choix d'annotation : le système des traits 5
 - 2.5 Cas controversés : les traits de l'unité <incertain> 7
 - 2.6 Les choix d'annotation : le système des valeurs 8
 - 2.7 Cas particuliers 8
- 3 Partie II : pour une étude de l'orthographe appliquée 9
 - 3.1 Conventions et définitions 9
 - 3.2 L'annotation des items orthographiques dans le corpus ÉNEPS 9
 - 3.3 Les choix d'annotation : le système des unités 10
 - 3.4 Les choix d'annotation : le système des traits 10
 - 3.5 Cas particuliers : les traits de l'unité <extra> 12
 - 3.6 Les choix d'annotation : le système des valeurs 12
 - 3.7 Cas particuliers 12
- 4 Bibliographie 13

1 Présentation générale et objectifs du manuel d'annotation

Le Guide d'annotation du Corpus ÉNEPS a été développé dans le cadre de la thèse intitulée *Travailler les compétences rédactionnelles à l'ÉNEPS* (École Nationale de l'Enseignement Professionnel Supérieur). Il précise les objectifs et la méthodologie d'annotation d'un corpus d'écrits académiques non normés faisant partie du corpus *Littérature Avancée*, présent au LILILEM (Jacques & Rinck, 2017).

Le guide a comme but de décrire le système d'étiquetage mis au point pour le **Corpus ÉNEPS**, et de fournir les éléments de compréhension des choix théoriques faits avant et pendant le processus d'annotation. Il se compose de deux parties qui correspondent à deux modèles d'annotation distincts. Dans la partie I, on explicite le protocole d'annotation pour une étude des chaînes de référence ; dans la partie II, on détaille le protocole d'annotation pour une étude de l'orthographe appliquée (Pallanti, Brissaud & Jacques, 2020)¹.

1.1 Composition du corpus ÉNEPS

Les textes composant le **Corpus ÉNEPS** ont été recueillis durant l'expérimentation du module PADE-PRO (Parcours d'Apprentissage et Développement de l'Écrit Professionnel), de septembre 2018 à mars 2019. Les analyses développées sur le **Corpus ÉNEPS** s'inscrivent également dans le cadre du projet ANR **E-CALM**. Le corpus est constitué par de textes manuscrits. Sa taille relativement réduite nous permet d'envisager une annotation manuelle. Voici le détail de la composition globale du Corpus ÉNEPS.

Tableau 1 : Composition globale du **Corpus ÉNEPS**

Section du corpus	Nb textes	Nb approximatif mots	Nature textes	Normativité textes
Section Expérimentale (1/2)	98	23.070	manuscrite	non normés
Section Résumé (2/2)	140	14.400	manuscrite	non normés
Corpus ÉNEPS	238	37.470	manuscrite	non normés

Clé de lecture : Le **Corpus ÉNEPS**, composé de la section **Expérimentale** et de la section **Résumé**, comporte au total 238 textes manuscrits et environ 37.470 mots. Il appartient à la catégorie des corpus non normés.

Ce guide d'annotation ne concerne que la section **Expérimentale** du **Corpus ÉNEPS**.

1.2 Transcription et pré-annotation

L'importante présence de textes manuscrits et le travail de transcription nécessaire pour constituer un corpus numérique, nous a **obligés** à adopter une pré-annotation au moment de la transcription du corpus. Dans le tableau ci-dessous, on peut observer l'ensemble des choix relatifs à la pré-annotation.

¹ Article accepté, en cours de publication.

4. Méthodes d'analyse et résultats

4.1 Chaines de référence : analyse quantitative (a)

Résultats de l'annotation des référents

Tableau XXXIII - Résultats de l'annotation des référents pour les groupes *Expérimental* (N=36) et *Contraste* (N=13)

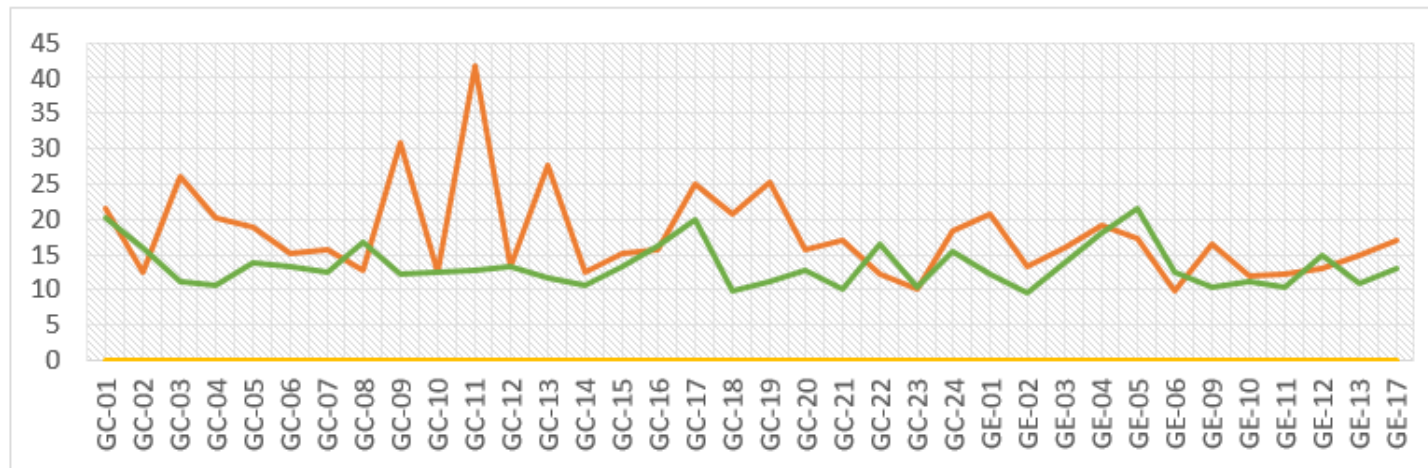
	Nombre moyen de mots, de maillons et de mots entre deux maillons dans les pré et post-tests					
	Nb moyen maillons		Nb moyen mots		Nb moyen mots/maillons	
	<i>Pré</i>	<i>Post</i>	<i>Pré</i>	<i>Post</i>	<i>Pré</i>	<i>Post</i>
Groupe <i>Expérimental</i>	12,22	18,28	198,06	235,89	17,83	13,36
Groupe <i>Contraste</i>	13,85	12,00	206,85	206,54	16,40	18,81

Clé de lecture : de gauche à droite, pour le groupe *Expérimental* le nombre moyen de maillons annotés par texte passe de 12,22 dans les pré-tests à 18,28 dans les post-tests, alors que le nombre moyen de mots par texte passe de 198,06 dans les pré-tests à 233,89 dans les post-tests. De même, le nombre moyen de mots entre deux maillons passe de 17,83 dans les pré-tests à 13,36 dans les post-tests (les tables des données complètes sont disponibles en annexe 9).

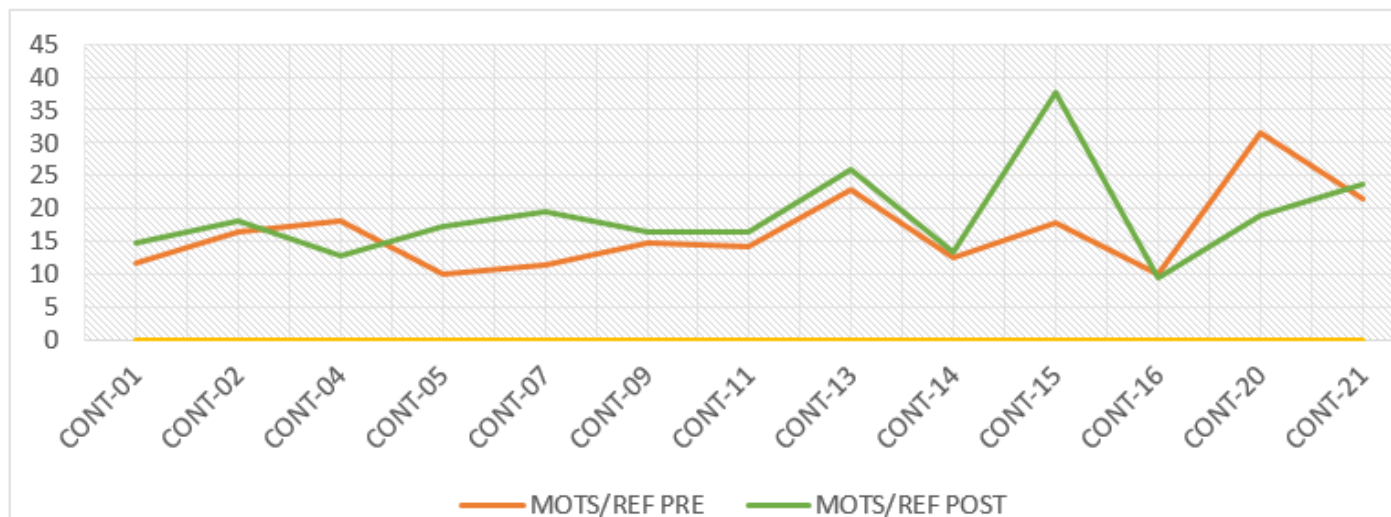
Test de Wilcoxon : p-value = ,000*

4.2 Chaines de référence : analyse quantitative (b)

Nombre moyen de mots entre deux maillons



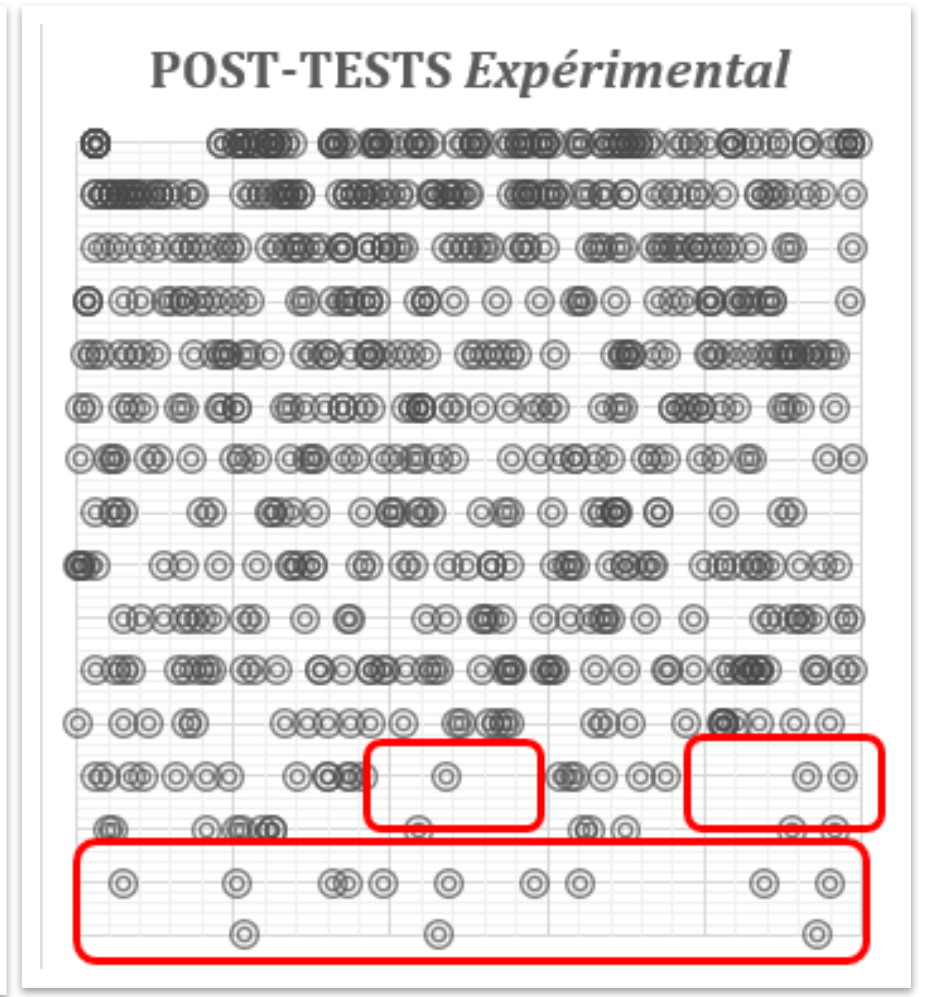
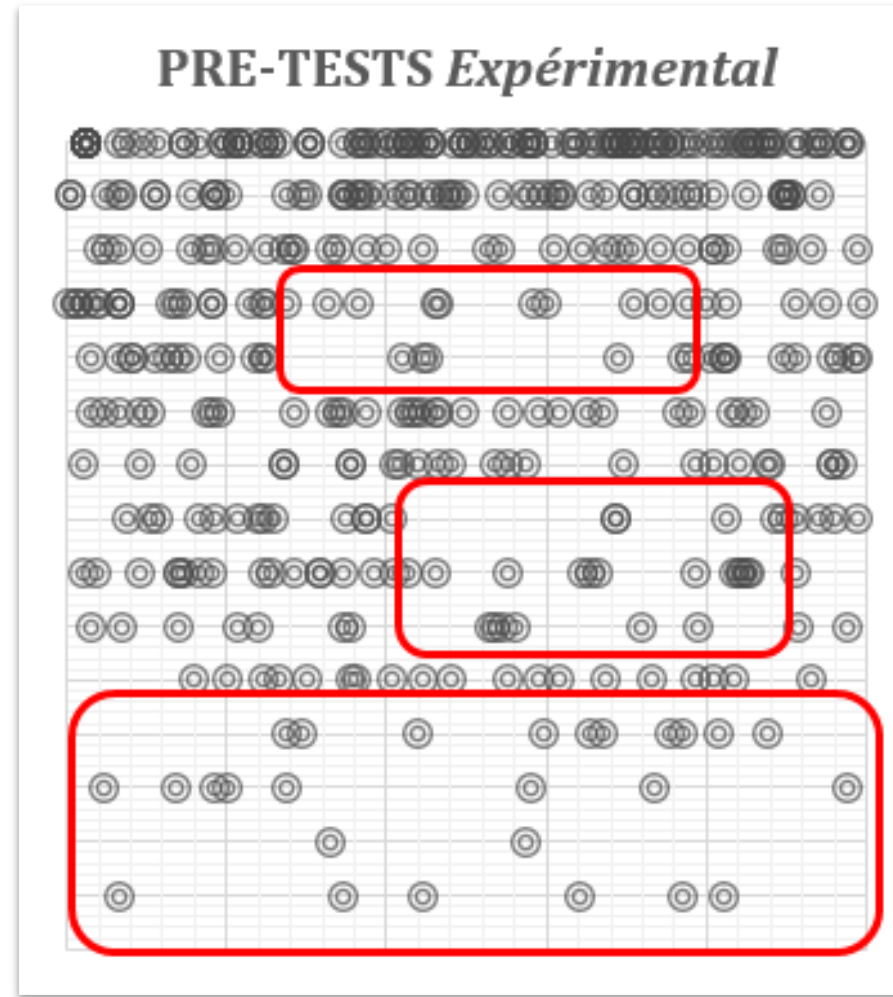
groupe *Expérimental*
(N=36)



groupe *Contraste*
(N=13)

4.3 Chaines de référence : analyse quantitative (c)

Distribution
des référents
dans les textes



4.4 Exemple : un cas

Reproductions des textes réalisés par l'étudiant GE-1 lors du pré et du post-test

Pré-test

Alain défend la position qu'il faut encouragé l'économie de partage mais avec certaines limites tel que de ne pas faire de concurrence déloyale à des entreprises qui sont-elles imposable et plus cher tel que Uber qui faisait de l'ombre au taxis en vu de ces faibles coup de transport.

Afin de défendre sa position Alain utilise divers arguments qui sont les suivants: Tout d'abord il évoque le fait que l'économie de partage c'est largement développée ces dernières années autour des plateformes internet, ce qui rend donc cela rapide et simple d'utilisation. Il parle notamment de covoiturage qui est très avantageux pour les clients de ce système de partage tout comme pour les personnes proposant le service. En vue du faite que les personnesp roposant le trajet le font déjà de base elle ne fond que rentabilisé une partie de celui-ci et de la même manière elle propose une service à un coup réduit au clients. Il évoque aussi RBNB qui est un site permettant de proposé à un particulier de louez leur maison durant une période déterminé où les loueurs n'y seraient pas. Cela montre donc bien que RBNB, ou d'autres site de partages, permet simplement de rentabilisé un bien ou un service dont nous n'avont de toutes manière pas besoin ou que l'ont fait déjà.

Post-test

Alain défend la position que mangé de la viande de temp en temp n'est pas mauvais, au contraire.

Pour défendre ce point de vue il va utilisé divers arguments pour répondre au sujet de certaines controverses sur la viande.

Tout d'abord il va évoqué la controverse de la santé et y répondre en mentionnant le fait que ne pas mangé de viande est mauvais pour la santé.

Il évoque le fait qu'une personne qu'il connaît à arrêté de mangé de la viande et quel s'est senti très fatigué au début, De plus il évoquera le faite que l'Homme est homnivores et qu'ils mangent de la viande depuis toujours Comme nous avons pue le constaté avec les peintures préhistorique.

En deuxieme, il évoquera l'écologie et le faite qu'il faut 1000L d'eau pour 1kg de boeuf. Il répondra à cela en affirmant que la terre entiere ne mange pas tous de la viande et que nous avons besoin de tout type de nourriture pour vivre correctement.

En dernier il évoquera la souffrance animal et utilisera le faite qu'il y a une grande zone d'ombre sur cela et qu'il est clair qu'il faudrait arrêté d'en mangé si les animaux subissaient une grande souffrance.

Il conclue en disant qu'il faut mangé de la viande mais pas trop en abusé.

4.5 Analyse de la compétence orthographique

Taux moyens de réussite dans les rubriques ORT-DICT et ORT-REDAC net variation absolue entre pré-tests et post-tests

	accAdj		conjVer		participePasse	
	Pré-test %	Post-test %	Pré-test %	Post-test %	Pré-test %	Post-test %
<i>Expérimental</i> ORT-DICT	41.67	58.89	47.35	50.97	49.49	45.37
Var. absolue %	+17,22		+3,62		-4,12	
<i>Expérimental</i> ORT-REDAC	71.10	76.34	85.25	88.75	54.22	62.95
Var. absolue %	+5,24		+3,5		+8,63	
<i>Contraste</i> ORT-DICT	25	47.69	46.15	51.15	39.86	46.15
Var. absolue %	+22,69		+5		+6,29	
<i>Contraste</i> ORT-REDAC	72.35	54.92	89.26	88.26	55	63.00
Var. absolue %	-17,43		-1		+8	

Clef de lecture : le groupe *Expérimental* affiche un taux moyen de réussite de 41,67% dans le pré-test et de 58,89% dans le post-test de la rubrique ORT-DICT, dans le secteur *accAdj*. La variation absolue entre pré-test et post-test correspond à +17,22%.

4.6 Analyse de la compétence orthographique

Résultats dans la catégorie accord verbal

Tableau XLII - Résultats du groupe *Expérimental* dans la catégorie *conjVer* dans la rubrique ORT-REDAC (N=36)

lemme	nb Pré	nb err pré	% réussite	nb post	nb err post	% réussite
être	152	14	90,79	258	13	94,96
avoir	51	19	62,75	101	21	79,21
falloir	28	2	92,86	103	1	99,03
-er (sauf aller)	222	43	80,63	301	47	84,39
autres	231	23	90,04	175	27	84,57

groupe *Expérimental*
(N=36)

Tableau XLIII - Résultats du groupe *Contraste* dans la catégorie *conjVer* dans la rubrique ORT-REDAC (N=13)

lemme	nb pré	nb err pré	% réussite	nb post	nb err post	% réussite
être	70	5	92,86	64	2	96,88
avoir	16	3	81,25	31	4	87,10
falloir	30	0	100,00	8	1	87,50
er (sauf aller)	86	15	82,56	96	21	78,13
autres	81	10	87,65	95	6	93,68

groupe *Contraste*
(N=13)

5. Conclusion et perspectives

5.1 Conclusion : évaluer la compétence textuelle

Compétences textuelles

L'annotation des référents rend possible une approche de la compétence textuelle

→ **une gestion correcte des référents participe à la bonne formation textuelle et à son intelligibilité.**

Améliorations

À développer

- Harmonisation finale des choix d'annotation (adjudication)
- Harmonisation du codage
- Révision finale du manuel d'annotation

- Analyse croisée des guides d'annotation E:CALM pour aboutir à un guide exploitable à des fins didactiques
- Développement de méthodes d'analyse quantitative de la coréférence

5.2 L'apport de la linguistique de corpus outillée

Le motif essentiel de la linguistique de corpus est à notre avis l'intérêt grandissant de la linguistique pour **ces aspects impossibles à traiter par l'introspection et l'intuition**. Il ne semble plus satisfaisant de s'intéresser à la langue dans ce qu'elle a de systématique sans rendre compte en même temps des situations réelles dans lesquelles les règles ainsi élaborées se réalisent, c'est-à-dire sans déterminer la sphère d'application de ces règles et leurs conditions de variation. **La linguistique de corpus apparaît bien alors comme le meilleur moyen d'atteindre la diversité des faits de langue authentiques.**

(Jacques, 2005, p. 24)



5.3 Evaluer des dispositifs didactiques expérimentaux

Mise au point d'un faisceau d'indicateurs de réussite pour évaluer les compétences rédactionnelles en production d'écrit.



Méthodes de la linguistique de corpus outillée



Elaboration d'outils d'évaluation des dispositifs didactiques expérimentaux visant l'amélioration des compétences rédactionnelles ?

Pour ne pas conclure...



« Depuis environ 15 ans, on assiste à l'expansion de sociétés privées (*Projet Voltaire, Orthodidacte*) sur le marché de l'éducation, publique et privée ; dans la plupart des cas, ces sociétés proposent des dispositifs d'évaluation automatique standardisée qui ne s'appuient pas sur des bases scientifiques solides. Nous pensons alors que le moment est venu pour les institutions scientifiques publiques de reprendre la main sur une compétence aussi importante que la compétence rédactionnelle, dans le but de développer des outils à même de garantir une évaluation équitable et scientifiquement valide des capacités à l'écrit des étudiants français. »

Pallanti, 2021

6. Bibliographie

- Bloomfield, L. (1925). Why a linguistic society? *Language*, 1, 1-5.
- Boch, F., Sorba, J., & Bessonneau, P. (2016). Évaluer les compétences rédactionnelles : Que tester ? *Le français aujourd'hui*, 193, 127-144. <https://doi.org/10.3917/lfa.193.0127>
- Charolles, M. (1978). Introduction aux problèmes de la cohérence des textes. *Langue française*, 38(1), 7-41. <https://doi.org/10.3406/lfr.1978.6117>
- Charolles, M. (2002). *La référence et les expressions référentielles en français*. Editions Ophris.
- Elalouf, M.-L., & Boré, C. (2007). Construction et exploitation de corpus d'écrits scolaires. *Revue française de linguistique appliquée*, Vol. XII(1), 53-70.
- Harris, Z. S. (1969). Analyse du discours (F. Dubois-Charlier, Trad.). *Langages*, 4(13), 8-45. <https://doi.org/10.3406/lgge.1969.2507>
- Jacques, M.-P. (2005). Pourquoi une linguistique de corpus ? In G. Williams (Éd.), *La linguistique de corpus* (p. 21-30). Presses Universitaires de Rennes.
- Jacques, M.-P. (2016). Une linguistique outillée, pour quels objets ? *Histoire Epistémologie Langage*, 38(2), 87-99. <https://doi.org/10.1051/hel/2016380207>
- Pallanti, L. (2021). *Travailler les compétences rédactionnelles à l'ENEPS. Conception et mise en oeuvre d'un système didactique expérimental*. [Thèse de doctorat]. Univ. Grenoble Alpes.
- Pallanti, L., Brissaud, C., & Jacques, M.-P. (2019). Conception d'un parcours d'apprentissage de l'écrit dans les filières professionnelles universitaires. *Revue internationale Methodol. Méthodologie de l'enseignement et de l'apprentissage des langues*, 3. <https://methodal.net/?article238>
- Pallanti, L., Brissaud, C., & Jacques, M.-P. (2020). Evaluer l'évolution des performances orthographiques d'étudiants en formation professionnelle : Méthodes et perspectives. *SHS Web of Conferences. 7e Congrès Mondial de Linguistique Française*, 78, 07018. <https://doi.org/10.1051/shsconf/20207807018>
- Pallanti, L., Jacques, M.-P., & Brissaud, C. (2021). Travailler l'écrit pour favoriser la réussite des étudiants issus de bacs professionnels : Un enjeu de linguistique appliquée. *Études de linguistique appliquée*, 2(202), 167-179.
- Poudat, C., & Landragin, F. (2017). *Explorer un corpus textuel. Méthodes—Pratiques—Outils*. De Boeck Supérieur. <https://doi.org/10.3917/dbu.pouda.2017.01>
- Schnedecker, C. (2017). Les chaînes de référence : Une configuration d'indices pour distinguer et identifier les genres textuels. *Langue française*, 195, 53-72. <https://doi.org/10.3917/lf.195.0053>
- Wüster, E. (1931). *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik : Die nationale Sprachnormung und ihre Verallgemeinerung*. Bouvier.