

Formaliser le vocabulaire

max.silberztein@univ-fcomte.fr

Limiter la taille du vocabulaire ?

- Il est important de pouvoir limiter la taille du vocabulaire à un ensemble de taille finie... Mais est-ce possible ?

Limiter la taille du vocabulaire ?

- Il est important de pouvoir limiter la taille du vocabulaire à un ensemble de taille finie... Mais est-ce possible ?
- Que faire des vocabulaires de chacun ?

Limiter la taille du vocabulaire ?

- Il est important de pouvoir limiter la taille du vocabulaire à un ensemble de taille finie... Mais est-ce possible ?
- Que faire des vocabulaires de chacun ?
- Comment traiter l'évolution de la langue (langue *vivante*) ?

Limiter la taille du vocabulaire

- Les vocabulaires de spécialité :
 - Vocabulaires techniques : termes de plomberie, de cuisine, d'électricité, d'architecture, de comptabilité, informatique (ex. *un langage orienté objet*)

Limiter la taille du vocabulaire

- Les vocabulaires de spécialité :
 - Vocabulaires techniques : termes de plomberie, de cuisine, d'électricité, d'architecture, de comptabilité, informatique (ex. *un langage orienté objet*)
 - Vocabulaires scientifiques : astronomie, biologie moléculaire, physique des particules, optique des lasers, chimie (ex. *un adduit*)

Limiter la taille du vocabulaire

- Les vocabulaires de spécialité :
 - Vocabulaires techniques : termes de plomberie, de cuisine, d'électricité, d'architecture, de comptabilité, informatique (ex. *un langage orienté objet*)
 - Vocabulaires scientifiques : astronomie, biologie moléculaire, physique des particules, optique des lasers, chimie (ex. *un adduit*)
 - Vocabulaires des arts : théâtre, sculpture, peinture, opéra, cinéma, photographie, musique (ex. *une dominante*)

Limiter la taille du vocabulaire

- Les vocabulaires de spécialité :
 - Vocabulaires techniques : termes de plomberie, de cuisine, d'électricité, d'architecture, de comptabilité, informatique (ex. *un langage orienté objet*)
 - Vocabulaires scientifiques : astronomie, biologie moléculaire, physique des particules, optique des lasers, chimie (ex. *un adduit*)
 - Vocabulaires des arts : théâtre, sculpture, peinture, opéra, cinéma, photographie, musique (ex. *une dominante*)
 - Vocabulaires des jeux et loisirs : jeu d'échecs, de go, de dames, de tarot, de belotte, de poker, philatélie (ex. *un gaufrage*)

Limiter la taille du vocabulaire

- Les vocabulaires de spécialité :
 - Vocabulaires techniques : termes de plomberie, de cuisine, d'électricité, d'architecture, de comptabilité, informatique (ex. *un langage orienté objet*)
 - Vocabulaires scientifiques : astronomie, biologie moléculaire, physique des particules, optique des lasers, chimie (ex. *un adduit*)
 - Vocabulaires des arts : théâtre, sculpture, peinture, opéra, cinéma, photographie, musique (ex. *une dominante*)
 - Vocabulaires des jeux et loisirs : jeu d'échecs, de go, de dames, de tarot, de belotte, de poker, philatélie (ex. *un gaufrage*)

Les variantes régionales

- Le français de Suisse, de Belgique, du Canada, du Sénégal

Les variantes régionales

- Le français de Suisse, de Belgique, du Canada, du Sénégal
- Les termes utilisés en Picardie, en Franche-Comté, en Bretagne ou en Provence

Les variantes régionales ou sociales

- Le français de Suisse, de Belgique, du Canada, du Sénégal
- Les termes utilisés en Picardie, en Franche-Comté, en Bretagne ou en Provence
- Les termes "jeunes", de la banlieue, enfantins, etc. (ex. *Black English* aux Etats-Unis)

Quelques expressions utilisées dans l'hexagone

aimable comme une porte de prison (= désagréable)

cousu de fil blanc (= supercherie vite découverte)

être sur des charbons ardents (= attendre une nouvelle avec anxiété)

faut pas pousser mémé dans les orties (= il ne faut pas abuser)

ours mal léché (= personne impolie)

passer sous les fourches caudines (= subir une humiliation)

renvoyer aux calendes grecques (= ajourner à une date indéterminée)

sortir quelque chose de derrière les fagots (= retrouver quelque chose de précieux)

tirer les marrons du feu (= entreprendre une action pour le profit de quelqu'un d'autre, tirer avantage d'une situation)

ton père il n'est pas vitrier (= tu obstrues mon champ de vision)

Les variantes régionales ou sociales

- Le français de Suisse, de Belgique, du Canada, du Sénégal
- Les termes utilisés en Picardie, en Franche-Comté, en Bretagne ou en Provence
- L'argot, le verlan, les termes « jeunes », de la banlieue, enfantins, etc. (ex. *Black English* aux Etats-Unis)
- **Trouvez 10 termes ou expressions régionales**

Quelques expressions à Besançon

être gaugé (= être trempé)

ça meule (= il fait froid)

s'entrouper (= trébucher)

T'as meilleur temps de (= tu ferais mieux de)

le cheni (= la poussière)

le cornet (= sac plastique)

la clenche (= la poignée)

un beuillot (= un idiot)

beugner (= cogner)

Quelques expressions en Afrique

abacost = veston sans col (Burundi, RDC, Rwanda)

désarmer = supplanter quelqu'un dans une relation amoureuse (Centrafrique, Rwanda, Tchad)

indexer = montrer du doigt, juger (Centrafrique, Tchad)

jeton = fric (Centrafrique, Congo, Tchad)

kwashiorkor = carence en protéine, enfants (Centrafrique, Rwanda, Burundi)

lessiver = réprimander (Cameroun)

ouissam = médaille (Algérie, Maroc)

potopoto = boue (RDC, Rwanda)

ziboulateur = décapsulateur (Burundi, Congo, RDC)

Quelques expressions à Montpellier

être ensuqué (= être dans les vaps)

un pitchoun (= un enfant)

le cagnard (= le soleil qui tape fort)

ça pègue (= c'est collant, c'est moite)

s'espanter (= être choqué)

s'enfader (= s'énerver)

un fada (= un fou)

avoir la cagne (= avoir la flemme)

une poche (= sac plastique)

escagasser (= embêter quelqu'un très fort)

tarpin (= très)

se faire mancher (= avoir une contravention)

Argot et parler « jeune »

- **crari** (= semblant, ex. *Fais pas crari, t'as pas la ref !*)
- **pranker** (= faire une farce)
- **ma gow** (= ma petite amie)
- **s'enjailler** (= faire la fête)
- **un charo** (= un dragueur)
- **un debz** (= un fourbe, ex. *Méfie-toi, c'est un debz !*)
- **être shlag** (= être nul, ex. *Ce gars, il est shlag !*)
- **rodave** (= regarder, surveiller, ex. *Rodave le gars chelou !*)

Conclusion partielle

- Impossible de recenser tous les mots de tous les vocabulaires d'une langue
- On se contentera de décrire le vocabulaire standard, i.e. Le vocabulaire partagé par tous et utilisé par les média

2. Evolution de la langue

- Des termes sont abandonnés et deviennent obsolètes, ex. :

*démané, fazéièrent, fenestrales, flamberie, gerriera, inexpériente,
notaresse, regina, salice, terrestrement, vitichoura*

2. Evolution de la langue

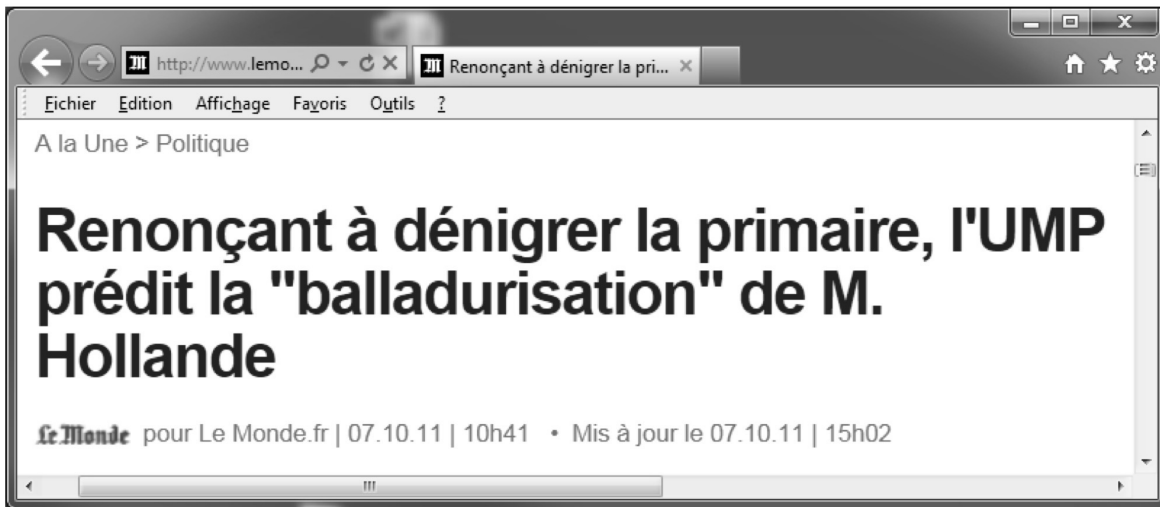
- Des termes sont abandonnés et deviennent obsolètes, ex. :

*démané, fazéièrent, fenestrales, flamberie, gerriera, inexpériente,
notaresse, regina, salice, terrestrement, vitichoura*

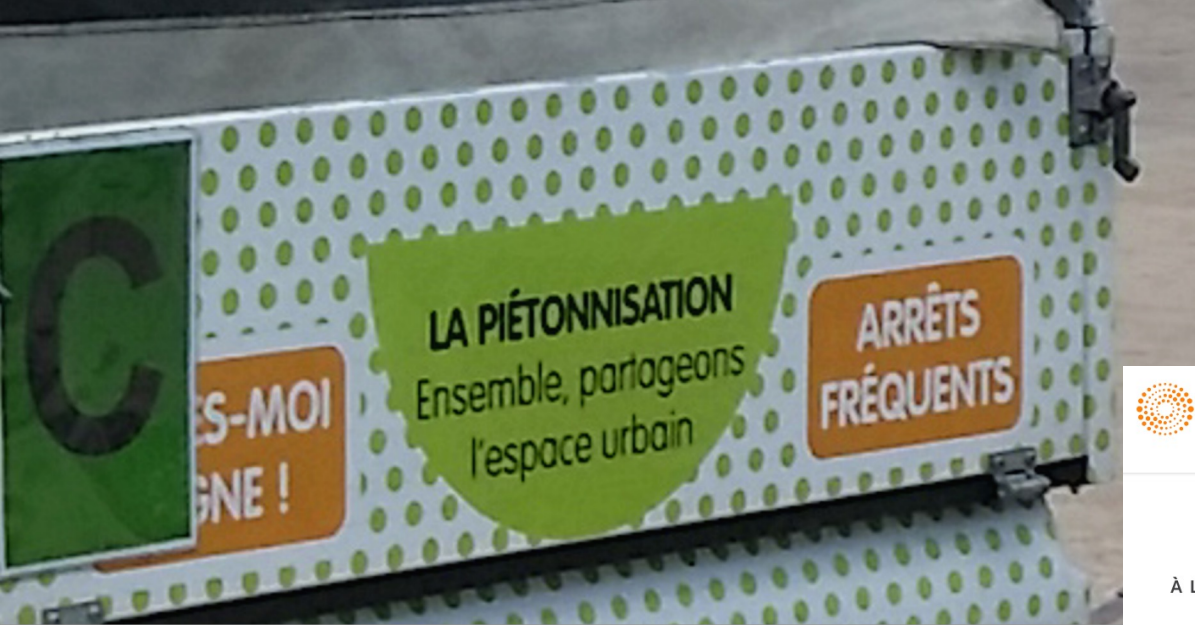
- De nouveaux termes sont créés tous les jours



une méga comparo, un FPS



... IL faut cependant rappeler que le prix demeure prohibitif eg qu'il existe des phablettes plus abordables (Le Monde 2014)



À Aix-en-Provence

À Paris :



À LA UNE 8 MARS 2018 / 08:41 / IL Y A UNE HEURE

Les Parisiens favorables à la piétonisation des voies sur berges, selon un sondage

Rédaction Reuters

2 MINUTES DE LECTURE



PARIS (Reuters) - Plus de la moitié des Parisiens (55%) se déclarent favorables au maintien de la piétonisation des berges de la rive droite de la Seine, d'après un sondage paru jeudi.





À Aix-en-Provence

À Paris :



Search...

À LA UNE 8 MARS 2018 / 08:41 / IL Y A UNE HEURE

Les Parisiens favorables à la piétonisation des voies sur berges, selon un sondage

Rédaction Reuters

2 MINUTES DE LECTURE



PARIS (Reuters) - Plus de la moitié des Parisiens (55%) se déclarent favorables au maintien de la piétonisation des berges de la rive droite de la Seine, d'après un sondage paru jeudi.



Nouveaux usages

- du coup = ainsi
- du coup = donc
- du coup = alors
- du coup = tout à coup
- du coup = soudain
- du coup = en conclusion
- du coup = ce qui fait que
- du coup = conséquemment
- ...

Argot et parler jeune

- crari
- pranker
- ma gow
- s'enjailler
- un charo
- un debz
- être shlag
- rodave

Emprunts à d'autres langues

- allemand (un bunker, un hamburger)
- anglais (le web, un email, googler)
- grec (téléphone, théâtre, psychologie, philosophie)
- hébreu (un abbé, abracadabra, une cabale)
- japonais (judo, manga, saké)
- nahuatl (tomate, chocolat, patate)
- russe (un bistrot)
- sanskrit (yoga)
- etc.

Constructions morphologiques

- Préfixation ou suffixation, ex. *la globalisation*
- Abréviation, ex. *une comparo, un ordi*
- Composition, ex. *une carte bleue, une pomme de terre*

A partir d'un nom propre

- une poubelle, le marxisme, gödéliser un problème mathématique
- Le sida, se pacser
- Le macronisme, la déstalinisation

Moscovici met en garde ceux qui veulent sarkoyser le parti (AFP, 2008)

Règles morphologiques ?

- Soit le verbe "monter"
- Question : ce verbe accepte-t-il le préfixe "dé", le suffixe "able" ?

Règles morphologiques ?

- Soit le verbe "monter"
- Question : ce verbe accepte-t-il le préfixe "dé", le suffixe "able" ?
- Ca dépend de son sens :
 - *Luc a monté son train électrique => Luc a démonté son train électrique, Ce train électrique est montable ; il est démontable*

Règles morphologiques ?

- Soit le verbe "monter"
- Question : ce verbe accepte-t-il le préfixe "dé", le suffixe "able" ?
- Ca dépend de son sens :
 - *Luc a monté son train électrique => Luc a démonté son train électrique, Ce train électrique est montable ; il est démontable*
 - *Luc est monté dans sa chambre => *Luc est démonté de sa chambre. Sa chambre est montable; elle est démontable*

Règles morphologiques ? NON

- Soit le verbe "monter"
- Question : ce verbe accepte-t-il le préfixe "dé", le suffixe "able" ?
- Ca dépend de son sens :
 - *Luc a monté son train électrique => Luc a démonté son train électrique, Ce train électrique est montable ; il est démontable*
 - *Luc est monté dans sa chambre => *Luc est démonté de sa chambre. *Sa chambre est montable; *elle est démontable*
 - *Luc remonte dans la chambre ; il remonte son train électrique.*
 - *Luc remonte sa montre ; cela remonte à dix ans*

Les dérivations sont imprévisibles

- Les propriétés doivent être examinées au coup par coup
- Deux verbes synonymes : mêmes conjugaisons, mêmes compléments, mêmes propriétés syntaxiques, etc.

Cette affaire concerne Luc = Cette affaire regarde Luc

Les dérivations sont imprévisibles

- Les propriétés doivent être examinées au coup par coup
- Deux verbes synonymes : mêmes conjugaisons, mêmes compléments, mêmes propriétés syntaxiques, etc.

Cette affaire concerne Luc = Cette affaire regarde Luc

*Luc est concerné par cette affaire = *Luc est regardé par cette affaire*

Les dérivations sont imprévisibles

- Les propriétés doivent être examinées au coup par coup
- Deux verbes synonymes : mêmes conjugaisons, mêmes compléments, mêmes propriétés syntaxiques, etc.

On peut rire de Luc = Luc est risible

Les dérivations sont imprévisibles

- Les propriétés doivent être examinées au coup par coup
- Deux verbes synonymes : mêmes conjugaisons, mêmes compléments, mêmes propriétés syntaxiques, etc.

On peut rire de Luc = Luc est risible

*On peut sourire de Luc = *Luc est sourisible*

Autonomie des mots dérivés

ensablement, globalisation, immobilisme

Autonomie des mots dérivés

ensablement, globalisation, immobilisme

ensablement : *amas de sable formé par un courant d'eau ou par le vent*
[source : Dictionnaire de l'Académie française, 1932, 1935]

globalisation : *terme de finance qui désigne la mise en place d'un marché de l'argent unifié au niveau planétaire, ou alors, terme politique qui désigne (de façon péjorative) l'unification des cultures entraînant la disparition des cultures locales*

immobilisme : *tendance à valoriser le statu quo ; aversion au changement*

Mots dérivés « potentiels »

citronisme, démodisation, enwebement

- Formes potentielles construites à partir des noms *citron*, *mode* (ou *mod*), *web*

Sens potentiels du mot *démodisation*

- *la mode* :

Période pendant laquelle un objet, un concept, un artiste, etc. perd de son succès auprès du public. Ou alors : perte de la puissance d'influence des phénomènes de mode auprès des personnes en fonction de leur âge. Ou alors : réduction du budget marketing alloué à la promotion d'un produit. Ou alors : politique de lutte contre les effets nocifs de la mode auprès de la jeunesse

Sens potentiels du mot *démodisation*

- *un mode musical :*

Libération de la musique contemporaine des contraintes liées aux deux modes musicaux majeur et mineur. Ou alors : cheminement qui conduit une mélodie à changer plusieurs fois de mode tout au long d'une pièce musicale

Sens potentiels du mot *démodisation*

- *un mode de fonctionnement :*

Unification en vue d'une simplification des différents modes de fonctionnement d'un outil mécanique ou électrique. Ou alors : automatisaion du changement de mode de fonctionnement d'un logiciel qui devient alors capable de s'adapter aux contraintes matérielles en temps réel (mémoire utilisable, bande passante du réseau, température du processeur, etc.)

Sens potentiels du mot *démodisation*

- *un mod* :

processus qui consiste à bloquer ou à enlever d'un logiciel toutes ses extensions (ses « mods »).

Sens potentiels du mot *démodisation*

- Activation d'un sens particulier :

Les égarements et les conséquences tragiques de la mode (dépenses excessives dues à l'obsolescence rapide des produits commerciaux ; faillite de nombreux distributeurs ne pouvant pas gérer leurs stocks d'invendus ; explosion de l'anorexie chez les adolescents) étaient devenues un tel problème de société qu'une politique de démodisation fut implémentée en 2040, en concertation avec les ministères concernés.

Conclusion partielle

- Impossible de prendre en compte les nouveaux mots
- Les règles morphologiques ne sont pas fiables
- Les formes dérivées morphologiquement ont des sens autonomes

=> Il faut recenser tous les mots qui ont un sens particulier, et ne pas prendre en compte les formes potentielles

Les Unités Linguistiques Atomiques (ALU)

- Les ALU constituent le vocabulaire standard d'une langue
- Les ALU sont nécessairement en nombre fini
- Les ALU ne sont pas analysables, même lorsqu'elles semblent pouvoir être décomposées en éléments morphologiques plus petits

Classification des ALU

- L'ordinateur ne traite que des séquences de bits, qui représentent des suites de caractères (lettres et séparateurs), i.e. a accès aux textes par l'orthographe
- La classification proposée est fondée sur l'orthographe, et est donc arbitraire et n'a pas forcément de sens d'un point de vue linguistique
- Elle a cependant l'avantage d'être simple, objective et donc reproductible.

4 classes d'ALU

(1) Les mots simples

- Les *mots simples* sont les ALU qui apparaissent dans les textes sous la forme de séquences de lettres entre deux séparateurs

Il mange une pomme => 4 mots simples

4 classes d'ALU

(2) Les affixes

- Les affixes (*préfixes, infixes ou suffixes*) sont les ALU qui apparaissent dans les textes sous la forme de séquences de lettres pas forcément entre deux séparateurs

La redéstructuration de la compagnie => 4 mots simples, 4 affixes (re, dé, structur, ation)

4 classes d'ALU

(3) Les mots composés

- Les *mots composés* sont les ALU qui apparaissent dans les textes sous la forme de séquences de lettres et de séparateurs

Il mange de plus en plus de pommes de terre => 2 mots simples et 2 mots composés

4 classes d'ALU

(4) Les expressions

- Toutes les autres ALU...

4 classes d'ALU

(4) Les expressions

- Les *expressions* sont les ALU qui apparaissent dans les textes sous la forme de séquences discontinues de lettres et de séparateurs

Il ne prend pas ce problème en compte => 3 mots simples, 2 expressions (ne ... pas et prendre ... en compte)

Types d'expressions

- Les expressions usuelles :

*Il faut **prendre** ce problème **en compte***

*Ce TD **aura** toujours **lieu** en C22*

*Il **est** souvent **en retard***

*On **ne** vous **fera pas part** de notre réponse*

Types d'expressions

- Les expressions idiomatiques :

Luc a pris la vache par les cornes

Luc a pris le taureau par la corne

Luc a pris le taureau par les cornes

Luc a pris ce gros boeuf par les cornes

Types d'expressions

- Les expressions idiomatiques :

~~Luc a pris la vache par les cornes~~

~~Luc a pris le taureau par la corne~~

Luc a pris le taureau par les cornes

~~Luc a pris ce gros boeuf par les cornes~~

Types d'expressions

- Les associations verbe support / nom prédicatif

*avoir du courage, donner une gifle, être en retard, faire une pause,
mettre à disposition, participer à une manifestation, prendre un repas*

Types d'expressions

- Les associations verbe support / nom prédicatif

avoir du courage, donner une gifle, être en retard, faire une pause, mettre à disposition, participer à une manifestation, prendre un repas

- Variabilité : *Il a un grand courage, elle lui a donné une sacrée gifle, il est très en retard, nous ferons une petite pause, il l'a mis à ma disposition exclusive, elle a participé à une énorme manifestation, il a pris son repas*

Types d'expressions

- Les associations verbe support / nom prédicatif

avoir du courage, donner une gifle, être en retard, faire une pause, mettre à disposition, participer à une manifestation, prendre un repas

- Variabilité : *Il a un grand courage, elle lui a donné une sacrée gifle, il est très en retard, nous ferons une petite pause, il l'a mis à ma disposition exclusive, elle a participé à une énorme manifestation, il a pris son repas*

- Effacement, variations : *Luc a du courage. Cela lui donne du courage. Il a perdu son courage. Il a retrouvé son courage.*

Luc est en retard. En retard, Luc s'est excusé. Cela l'a mis en retard. Il a du retard, etc.

Quelques conséquences désagréables

Aujourd'hui

Quelques conséquences désagréables

Aujourd'hui

*À l'amiable, arc-boutant, clopin-clopant, d'aucuns, jusqu'au boutisme,
parce que*

(un peu plus de 1.000 mots composés contiennent des mots non autonomes)

Conclusion

- Il est possible de représenter le vocabulaire d'une langue
- On doit se placer en synchronie (on ne décrit pas l'évolution de la langue)
- On se limite au vocabulaire standard (on ne décrit pas les vocabulaires de spécialité)
- On classe tous les éléments du vocabulaire (les ALU) en 4 classes : les mots simples, les affixes, les mots composés et les expressions.

Exercice

- Donner la liste des ALUs potentiellement présents dans la phrase suivante :

Marie est peut-être dans la rue du Commerce