

Polysémie et troncation des noms en -ion en français

Rémi Anselme

Université Paris-Diderot

Sous la direction de
Olivier Bonami et Heather Burnett

Sémantique pour les ressources en morphologie dérivationnelle

Toulouse, le 08 juillet 2019

Introduction

La troncation

- « *Procédé d'abrègement des mots polysyllabiques qui consiste à supprimer une ou plusieurs syllabes [...]* » — TLFi
- Opération morphophonologique : Base → Troncat
- Exemple : INTRODUCTION → INTRO

Les noms déverbaux en *-ion*

- Ambiguïté des noms en *-ion* entre plusieurs interprétations :
 - 1 Noms d'événements complexes
 - 2 Noms d'événements simples
 - 3 Noms résultatifs
- Ce sont des noms polysémiques.

Plan de la présentation

- Ce que dit la littérature sur la troncation
 - Kerleroux [1999]
- Étude sur la distribution des sens :
 - 1 Distribution des sens actionnels
 - 2 Entropie des distributions
 - 3 Comparaison avec les diminutifs
- Traitement de la polysémie par la troncation

La classification des nominalisations

- Pour décrire les contraintes sur la troncation des noms en *-ion*, Kerleroux [1999] s'appuie sur la classification des nominalisations due à Grimshaw [1990] (voir aussi Borer [2003], Roy and Soare [2013]).

Dénotation	Type de nom	Syntaxe
événement	N d'événement complexe <i>La manifestation fréquente de la vérité</i>	spéciale
	N d'événement simple <i>Les manifestation fréquentes des gilets jaunes</i>	ordinaire
objet	N résultatif <i>Les dissertations sont corrigées</i>	

- **Syntaxe spéciale** : article défini singulier, arguments obligatoires, etc.

La généralisation de Kerleroux

- Kerleroux [1999] : La troncation ne s'applique pas aux noms d'événements complexes.
 - Nom d'événement complexe :
 - *La manifestation de la vérité aura pris cinquante ans.*
 - * *La manif de la vérité aura pris cinquante ans.*
 - Nom d'événement simple :
 - *La manifestation des étudiants a duré longtemps.*
 - *La manif des étudiants a duré longtemps.*
 - Nom résultatif :
 - *On corrige les dissertations du baccalauréat.*
 - *On corrige les disserts du bac.*

Une deuxième généralisation

- Restriction/modification du domaine de référence (Raccah [2015])
- « Opération de singularisation des dénominations » (Kerleroux [1999])
 - COMPILATION → COMPIL : édition musicale
 - DEMONSTRATION → DEMO : production technologique
 - INTOXICATION → INTOX : domaine intellectuel
 - RÉDACTION → RÉDAC : exercice scolaire

Pourtant...



Le Toutos Nantais

@weberien44

Suivre



En réponse à @Virgile_Klamp @Infraraid et 3 autres

Vous m'excuserez, mais je préfère avoir les liens des études directement plutôt qu'une compil d'articles de presse.

13:06 - 28 mai 2019



1



Pourtant...



Marie Seners

@MarieSeners

Suivre



François Chérèque a prévenu Bernard Thibault qu'il quittera le carré de tête de la manif si le défilé tourne à la démo de force antisarko

06:00 - 25 avr. 2012

14 Retweets



1



14



Pourtant...



Linouche
@Linouche_SOA

Suivre



Bon c'est bien mignon les soirées etc mais quand tu finis dans les toilettes à 3h de matin en hyper ventilation et une intox alimentaire alors que tu boss le lendemain c'est pas si drôle

02:07 - 29 juin 2019

1 Retweet 3 J'aime



Pourtant...



(• ε •)

@Ninjawina

Suivre



@InesJett Je me lance dans la rédac de
l'article, gros blocage sur l'intro #lablague

14:04 - 13 mars 2012



1



Pourtant...



Titoutg
@Titoutg

Suivre

En réponse à @CRolandin

Pour la manif de la vérité je ne soutiens
personne je laisse la justice agir. C'est ce
que vous devriez faire en tant qu'elu

04:20 - 25 juil. 2016



corinne audouin ✓
@cocale

Suivre

En réponse à @Mic_k_a @PatriceReviron

Non sur le fond aussi c'est n'importe quoi. en
quoi ça contribue à la manif de la vérité?

04:22 - 27 avr. 2017



Distribution des sens

- Il ne semble pas y avoir de contraintes catégoriques contre la troncation des noms d'événements complexes ou l'utilisation de la troncation avec les différents sens de la base.
- Mais il y a peut-être des contraintes probabilistes de ce type.
- Deux questions :
 - 1 Les troncats ont-ils plus volontiers un sens actionnel que leur base ?
 - 2 Les troncats sont-ils normalement sémantiquement plus spécialisés que leurs bases ?

Étude sur la distribution des sens

Méthodologie

- Elaboration d'une liste de 10 paires de lexèmes BASE/TRONCAT
- A partir du Wiktionnaire et du TLFi : Elaboration de listes de sens
- Intersection de listes de sens de manière à simplifier les sens possibles

Corpus : FRCOW (Schäfer [2015], Schäfer and Bildhauer [2012])

2016 version of the French web corpus
by COW - Corpora from the Web

- Corpus entier : 8 789 873 699 mots
- France, Après 1995, Forum et Non-Forum
- Tirage aléatoire d'exemples → Annotation manuelle de 2000 exemples

Exemple I

PROMOTION → PROMO

- 1 Action d'élever simultanément plusieurs personnes à un même grade.
 - [...] i.e. les *incompétences* et la **promotion** de généraux ayant faillis ce genre de choses que je trouve plus intéressant à débattre [...]
- 2 Ensemble des candidats entrés la même année dans certaines grandes écoles.
 - [...] Jean-Pierre Liégeois qui m'invite à présenter mon travail devant une **promotion** des élèves de L' ENA [...]
- 3 Nomination, accession à un grade, une dignité, une fonction supérieure.
 - [...] entrée dans le métier , formation carrière , salaires , **promotion** horaires , temps et conditions de travail [...]
- 4 Accession d'une personne, d'un groupe social à de meilleures conditions de vie.

Exemple II

- *La classe préparatoire TSI , un exemple de **promotion** sociale [...]*
- 5 Ensemble des moyens mis en œuvre pour promouvoir un bien, un service ou un événement.
 - *Il ne s'agit pas ici de faire la **promotion** du naturisme et de ses valeurs [...]*
- 6 Action de faire construire des immeubles en vue de les vendre ou de les louer.
 - *[...] acteur majeur de la **promotion** immobilière du Grand Ouest [...]*
- 7 Opération commerciale afin d'attirer l'attention sur une offre ou un produit.
 - *[...] Profitez des **promotion** pour acheter d'importantes quantités que vous congèlerez. [...]*

Résultats de l'annotation

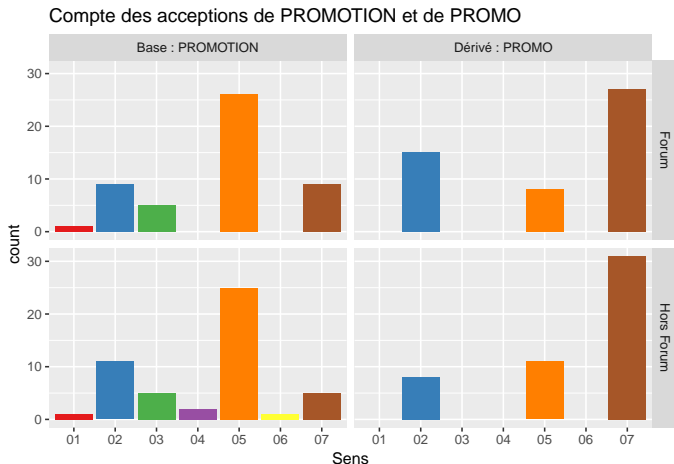


Figure 1 – Distribution des sens de PROMOTION/PROMO

Sens actionnel vs. non-actionnel

- Comme les noms d'événements complexes ne sont pas troncables, alors nous nous attendions à ce que les noms actionnels soient défavorisés.

	Non-forum		Forum	
	Sens non-actionnel	Sens actionnel	Sens non-actionnel	Sens actionnel
Base	269	181	279	171
Troncat	363	86	360	86

Table 1 – Compte des sens actionnels et non-actionnels

Sens actionnel vs. non-actionnel - Régression linéaire¹

- Est-il possible de prédire la forme d'un mot en fonction de son type de sens ?
- Ce modèle statistique permet de prendre en compte l'effet de chaque paire de lexèmes.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Troncat ~ Sens * Forum + (Sens * Forum | Lexeme)
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.05402    0.14055   0.384   0.701
## Sens         -1.45833    0.30797  -4.735 2.19e-06 ***
## Forum         0.06875    0.13449   0.511   0.609
## Sens:Forum   -0.29786    0.46132  -0.646   0.518
## ---
## Signif. codes:  0 '***' x 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Je remercie Doriane Gras qui m'a aidé à faire ce modèle.

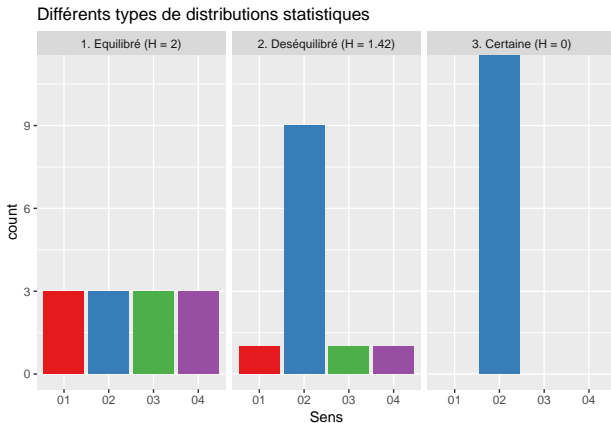
Entropie

L'entropie (Shannon [1948]) → rend compte de la quantité d'information “produite” :

$$H = \sum_{i=1}^n (p_i \times \log_2 p_i)$$

- Entropie nulle → Pas d'incertitude
- Entropie basse → Distribution statistique déséquilibrée ou/et avec peu de valeurs
- Entropie haute → Distribution statistique équilibrée mais beaucoup de valeurs possibles

Entropie - exemples



⇒ Entropie basse = un/quelques sens dominant(s).

Entropie

	Hors Forum		Forum	
	Base	Troncat	Base	Troncat
compilation	1.0512	0.5411	1.2299	0.9161
décoration	1.2597	0.8994	0.9284	0.9849
démonstration	1.9582	1.6319	2.2579	1.4325
information	2.3412	2.0024	2.0256	2.0995
intoxication	1.1273	1.1914	0.6395	1.1069
manifestation	1.6146	0.0000	2.1484	0.2422
promotion	2.0564	1.3311	1.8262	1.4241
récréation	1.6494	0.6800	1.0844	0.4021
rédaction	1.3925	1.1024	1.8111	0.8841
sonorisation	1.1974	0.3825	0.9593	0.8176

Table 2 – Mesure d'entropie pour les distributions de sens

Hors forum : $t(9) = -2.1741, p = 0.028$

Forum : $t(9) = -3.976, p = 0.001$

Conclusion intermédiaire

- 1 Les troncats ont plus facilement un sens actionnel que leur base
 - Cohérent avec l'hypothèse selon laquelle les emplois « Nom d'événement complexe » sont défavorisés par la troncation.
- 2 La distribution statistique des troncats est moins équilibrée que celle de leurs bases
 - Cohérent avec l'hypothèse que la troncation produit de la spécialisation sémantique (partielle).

Comparaison avec les verbes diminutifs

Pourquoi étudier les diminutifs ?

- *Machouiller, Pleuvoter, etc.*
- Comme pour la troncation, pas de changement dans le domaine de dénotation : mêmes sens que la base avec une distribution légèrement différente.
- Contrairement à la troncation, formés par suffixation et non par substraction de forme.

Utilisation de la même méthodologie :

- 1 Sélection de 5 paires de bases/dérivés
- 2 Élaboration de listes de définition (TLFI et Wiktionnaire)
- 3 Annotation manuelle

Exemple I

MÂCHER → MÂCHOUILLER

- 1 Action de broyer un aliment à l'aide des dents et de la mâchoire.
 - *Les oiseaux ne **mâchent** pas ce qu' ils mangent [...]*
- 2 Réduire en mordillant avec les dents et/ou en humectant de salive.
 - *[...] les feuilles de coca par exemple , **mâchées** dans certaines régions pauvres [...]*
- 3 Mordiller, mâchonner.
 - *[...] on ne me suce plus , on me **mâche** [...]*
- 4 Préparer pour faciliter la tâche à quelqu'un.
 - *[...] cherche tu trouveras je ne vais pas te **mâcher** le travail [...]*
- 5 Prononcer indistinctement.

Exemple II

- [...] et je **mâche** mes mots [...]
- 6 S'exprimer de façon claire et sans ménagement [si forme négative].
 - [...] Jean-Louis Triaud n' a pas **mâché** ses mots au micro [...]
- 7 Penser sans cesse à quelque chose.
 - Enfermé dans ses ratiocinations **mâchouillant** de mornes pensées [...]
- 8 Couper quelque chose sans faire une section nette, déchirer.
 - [...] on **mâche** dans la mouillure de la feuille de salade [...]
- 9 User ou marquer de façon irrégulière.
 - [...] il s' agit peut-être d' un docu déjà **mâché** [...]
- 10 Meurtrir.
 - Les combos lourds pour **mâcher** et broyer ses adversaires et les combos légers [...]

Résultats

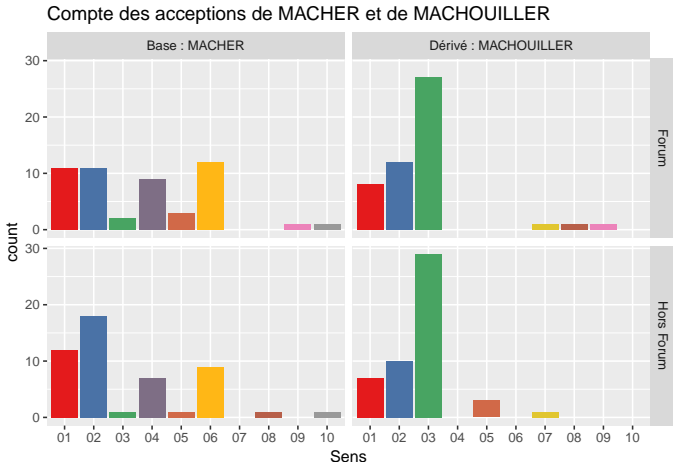


Figure 2 – Distribution des sens de MÂCHER/MÂCHOUILLER

Entropie des diminutifs

	Non-Forum		Forum	
	Base	Dérivé	Base	Dérivé
glander	0.9004	0.4689	1.0361	0.5411
gratter	3.0130	2.7911	3.2338	2.6688
mâcher	2.3186	1.6737	2.5556	1.7358
mordre	3.0859	0.7011	2.9665	0.5221
sauter	3.9235	2.3815	3.9431	1.9209

Table 3 – Mesure d'entropie pour les distributions de sens

No – forum : $t(4) = -2.59, p = 0.03034$

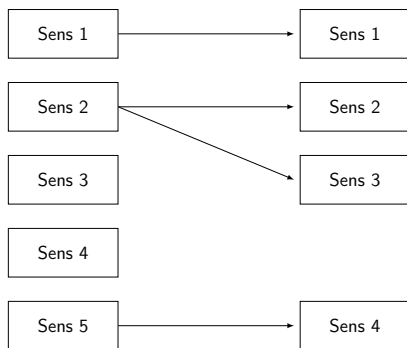
Forum : $t(4) = -3.1509, p = 0.01724$

Conclusion intermédiaire

- On observe une spécialisation partielle du sens des verbes diminutifs par rapport à leur base.
- Même résultat que pour la troncation.
- Suggère que cette propriété est générale pour les procédés de construction de lexème qui ne changent pas le domaine de dénotation.
- Est-ce que cela s'applique aussi aux autres procédés ?
- Difficile à établir : interaction entre polysémie de la base et polysémie du dérivé.
 - *batteur*₁ : machine utilisée en filature
 - *batteur*₂ : ustensile de cuisine
 - *batteur*₃ : ouvrier qui bat certaines matières
 - *batteur*₄ : musicien jouant de la batterie

Vers une approche probabiliste

Nous avons l'habitude de réfléchir à la polysémie en morphologie dérivationnelle en termes de relations catégorique entre sens de la base et sens du dérivé (voir p. ex. Fradin and Kerleroux [2009]).



Vers une approche probabiliste

Dans cette présentation nous avons montré des données fréquentielles qui suggèrent qu'on rate quelque chose d'important avec une telle approche.

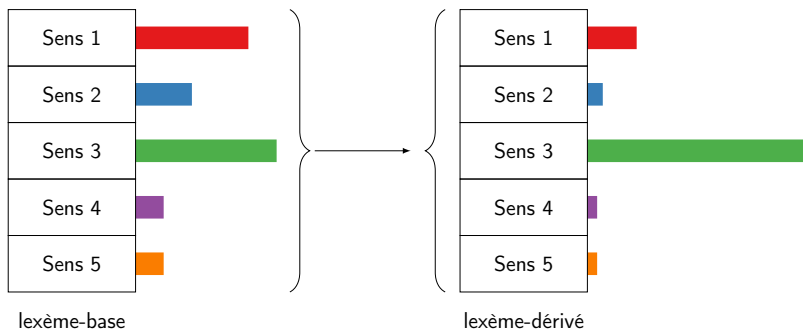
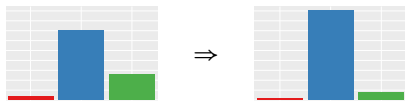


Figure 3 – RCL et distribution de probabilités des sens

Vers une approche probabiliste

- Version probabiliste de la théorie de Kerleroux :

- 1 Nous connaissons la distribution des bases et des troncats :



- 2 Nous cherchons à estimer x et y à partir des données :

- $x = P(\text{forme=troncat} \mid \text{sens=évt})$
- $y = P(\text{forme=troncat} \mid \text{sens=objct})$

- 3 Si le choix optimal de x et y est tel que $x \ll y$, alors la théorie de Kerleroux est confirmée.

- Dans ce modèle, il n'y a que deux facteurs qui influencent les distributions :

- 1 La distribution inhérente des sens à véhiculer, indépendamment du choix d'une forme pleine ou tronquée.
- 2 La préférence de chaque type de sens à être tronqué.

- Quels autres facteurs entrent en jeu ?

Conclusion générale

- Méthode pour explorer la polysémie des procédés constructionnels
 - Sélection d'un échantillon aléatoire d'attestations de bases et de dérivés.
 - Annotation des sens lexicaux rencontrés.
 - Analyse statistique de la distribution des sens rencontrés.
- Utilité :
 - Validation de contraintes non-catégoriques sur les procédés
 - Point de départ pour une modélisation probabiliste de la construction de lexèmes
- Limite : Annotation extrêmement coûteuse et bruitée.
- Des idées pour remplacer cette annotation manuelle ?

Merci.
Des questions ?

Références I

Françoise Kerleroux. Sur quelles bases opère l'apocope? Silexicales 2 : la morphologie des dérivés évaluatifs, pages 95–106, 1999.

Jane Grimshaw. Argument structure. the MIT Press, 1990.

Hagit Borer. Exo-skeletal vs. endo-skeletal explanations : Syntactic projections and the lexicon. The nature of explanation in linguistic theory, pages 31–67, 2003.

Isabelle Roy and Elena Soare. Event-related nominalizations. In Categorization and Cateogory Change, pages 123–152. Cambridge Scholars Publishing, Cambridge, 2013.

Pierre-Yves Raccah. Quelques infos sur les nœuds-pap et sur les collabos, 2015.

Références II

- Roland Schäfer. Processing and querying large web corpora with the COW14 architecture. In Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3), Lancaster, 2015. UCREL, IDS. URL <http://rolandschaefer.net/?p=749>.
- Roland Schäfer and Felix Bildhauer. Building large corpora from the web using a new efficient tool chain. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), pages 486–493, Istanbul, Turkey, 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL <http://rolandschaefer.net/?p=70>.
- Claude Elwood Shannon. A mathematical theory of communication. Bell system technical journal, 27(3) :379–423, 1948.
- Bernard Fradin and Françoise Kerleroux. L'identité lexémique, 2009.

Listes de lexèmes

compilation → compil
décoration → déco
démonstration → démo
information → info
intoxication → intox

manifestation → manif
promotion → promo
récréation → récré
rédaction → rédac
sonorisation → sono

glander → glandouiller
gratter → gratouiller
mâcher → mâchouiller
mordre → mordiller
sauter → sautiller

Troncation et indépendance des distributions

	Effet de Forum/Non-Forum						Effet de Troncat/Non-Troncat					
	Troncat			Non-Troncat			Forum			Non-Forum		
	χ^2	DL	Pval	χ^2	DL	Pval	χ^2	DL	Pval	χ^2	DL	Pval
Décoration	5.3752	1	0.0039	5.3752	2	0.0699	15,316	2	0.0009	5.0571	2	0.0599
Compilation	3.8655	2	0.1778	1.569	2	0.5485	11.663	2	0.0049	6.348	2	0.0229
Intoxication	2.0833	2	0.4186	3.4572	3	0.3397	74.785	2	0.0009	60.6	3	0.0009
Démonstration	5.6579	6	0.4575	6.3039	5	0.2957	60.202	7	0.0009	49.334	6	0.0009
Manifestation	2.0408	1	0.4955	20.057	5	0.0019	29.6	5	0.0009	47.059	3	0.0009
Récréation	2.2104	1	0.2318	8.2101	3	0.0339	6.5473	3	0.0359	15.082	3	0.0019
Sonorisation	3.5202	2	0.2118	4.2605	2	0.1159	34.41	2	0.0009	29.963	2	0.0009
Promotion	2.88	2	0.2517	4.3625	6	0.7223	26.029	4	0.0009	33.696	6	0.0009
Rédaction	3.4667	3	0.3626	14.466	4	0.0029	18.108	4	0.0009	36.641	4	0.0009
Information	5.7222	4	0.2158	9.8772	5	0.0669	1.1151	4	0.9061	20.036	5	0.0009

Table 4 – Résultats des test d'indépendance du χ^2

Diminutifs et indépendance des distributions

	Effet de Forum/Non-Forum						Effet de Dérivé/Non-Dérivé					
	Dérivé			Non-Dérivé			Forum			Non-Forum		
	Chi2	DL	Pval	Chi2	DL	Pval	Chi2	DL	Pval	Chi2	DL	Pval
Glander	1.1111	2	1	1.6123	3	0.7582	4.2941	3	0.2977	3.277	3	0.4126
Gratter	8.1208	9	0.5584	15.247	13	0.2787	30.292	13	0.0009	41.658	12	0.0009
Mâcher	5.3199	6	0.5165	5.745	8	0.7592	49.069	9	0.0009	49.735	8	0.0009
Mordre	6.3443	7	1	5.7808	11	0.9131	87.952	11	0.0009	70.328	11	0.0009
Sauter	20.978	9	0.0029	29.929	22	0.0649	73.938	19	0.0009	77.752	21	0.0009

Table 5 – Résultats des test d'indépendance du χ^2

Résultats



Figure 4 – Distribution des sens de COMPILATION/COMPIL

Résultats

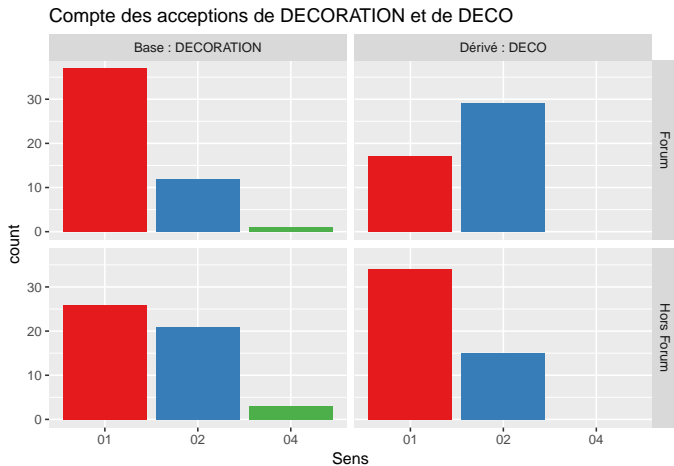


Figure 5 – Distribution des sens de DÉCORATION/DÉCO

Résultats

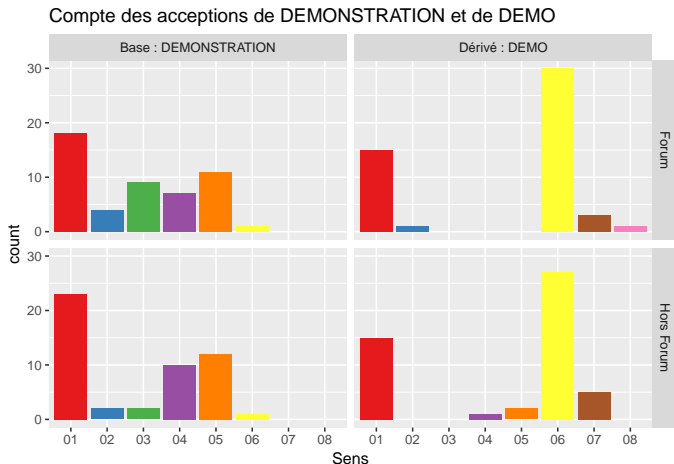


Figure 6 – Distribution des sens de DÉMONSTRATION/DÉMO

Résultats

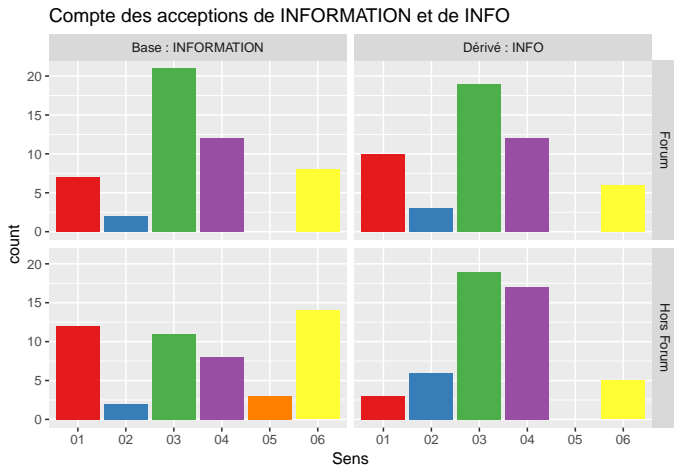


Figure 7 – Distribution des sens de INFORMATION/INFO

Résultats

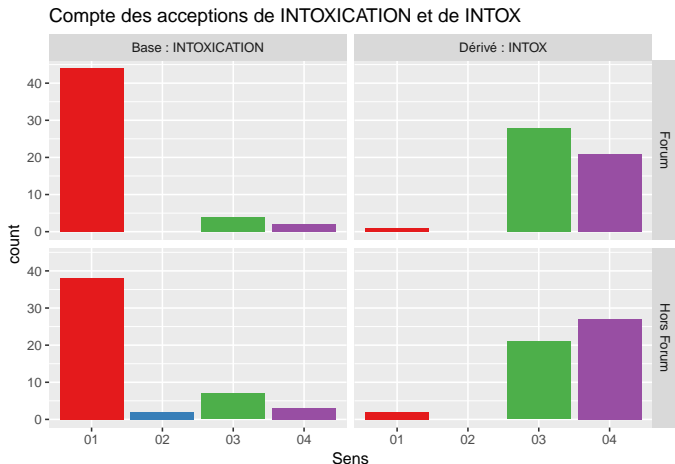


Figure 8 – Distribution des sens de INTOXICATION/INTOX

Résultats

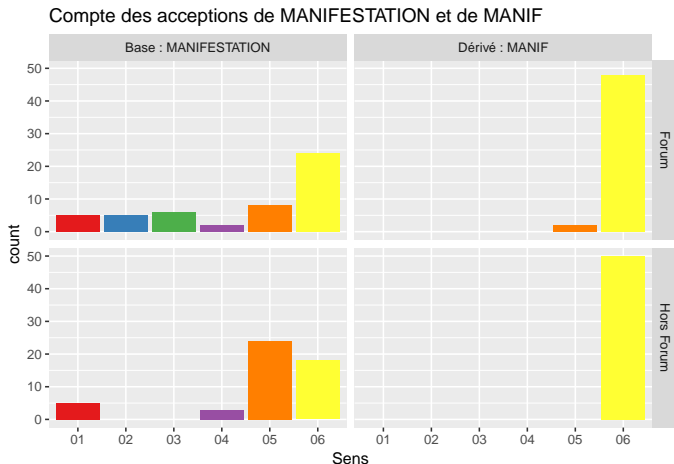


Figure 9 – Distribution des sens de MANIFESTATION/MANIF

Résultats

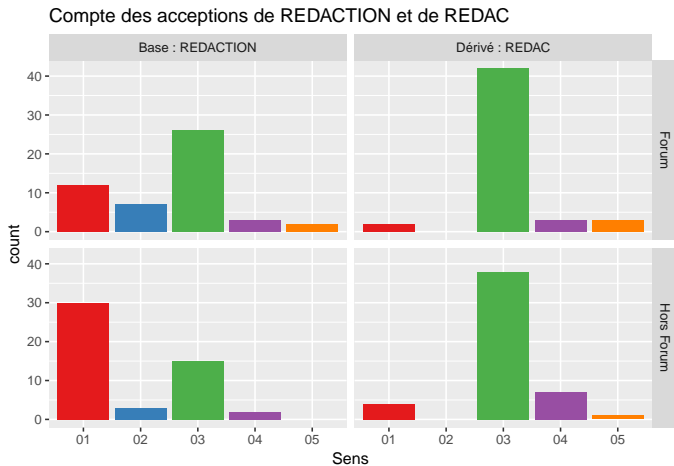


Figure 10 – Distribution des sens de RÉDACTION/RÉDAC

Résultats

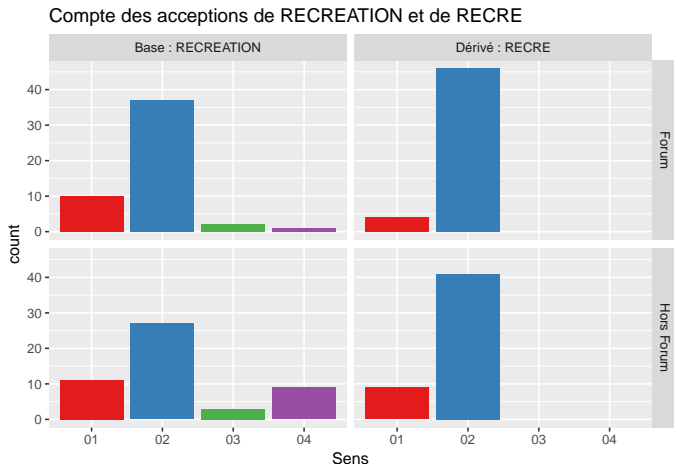


Figure 11 – Distribution des sens de RÉCRÉATION/RÉCRÉ

Résultats

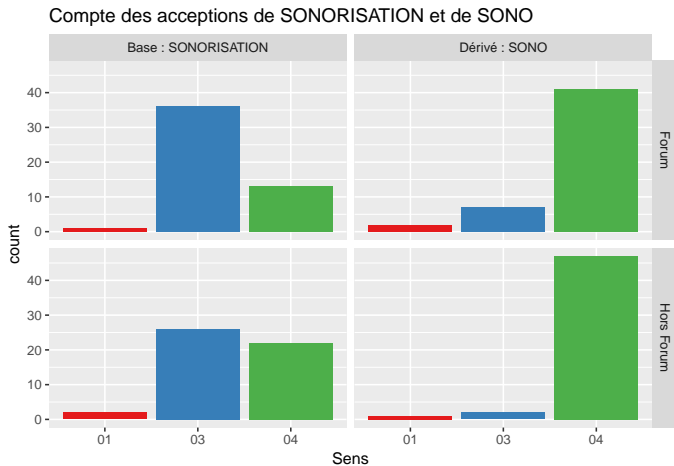


Figure 12 – Distribution des sens de SONORISATION/SONO

Résultats

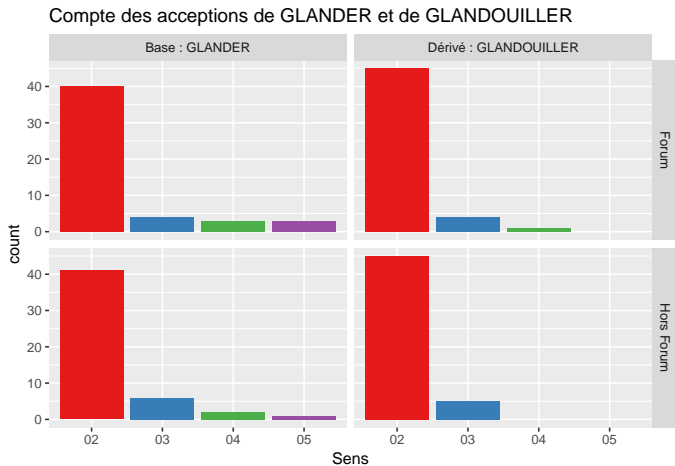


Figure 13 – Distribution des sens de GLANDER/GLANDOUILLER

Résultats

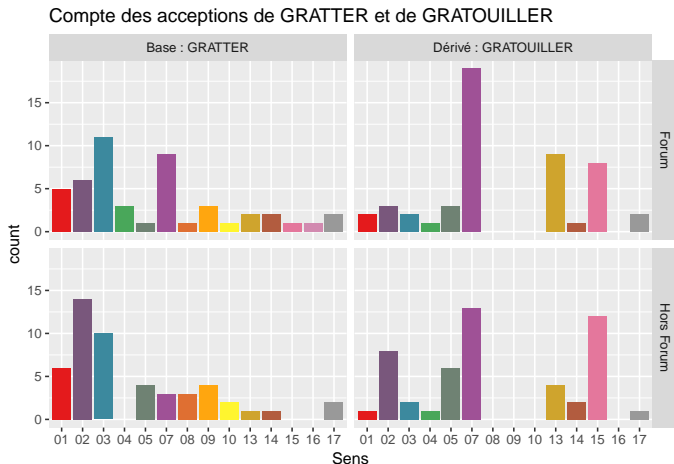


Figure 14 – Distribution des sens de GRATTER/GRATOUILLER

Résultats

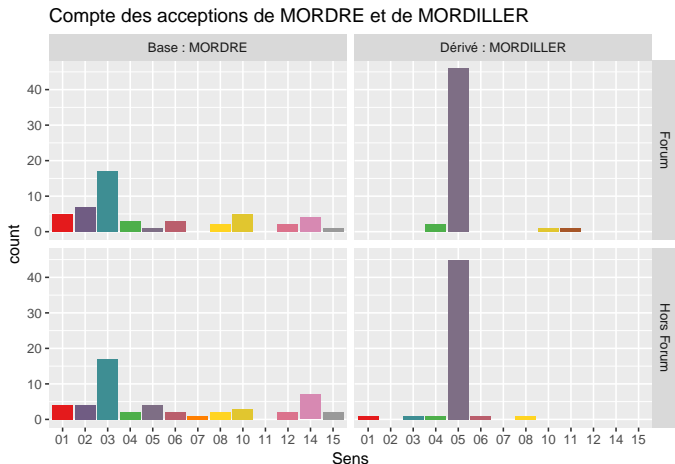


Figure 15 – Distribution des sens de MORDRE/MORDILLER

Résultats

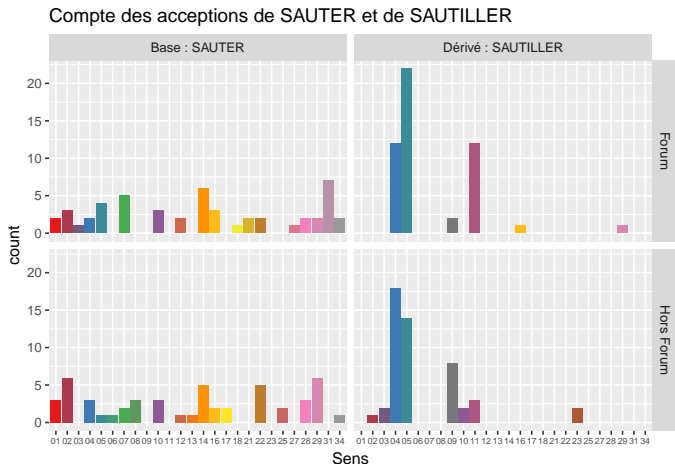
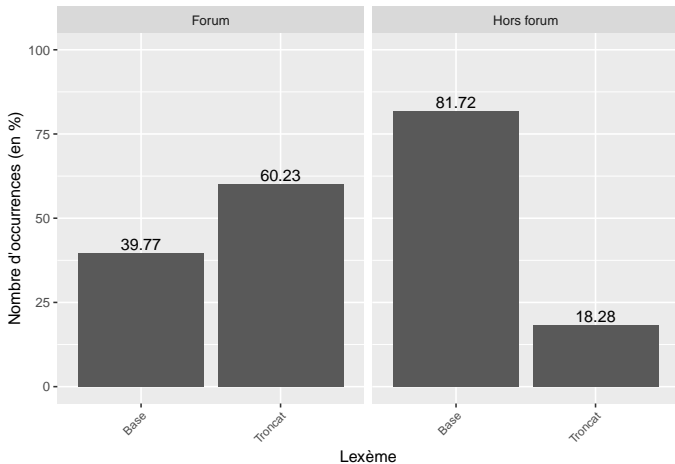
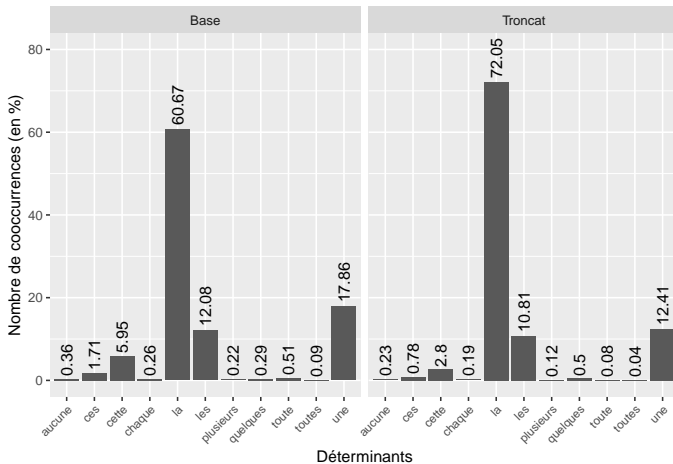


Figure 16 – Distribution des sens de SAUTER/SAUTILLER

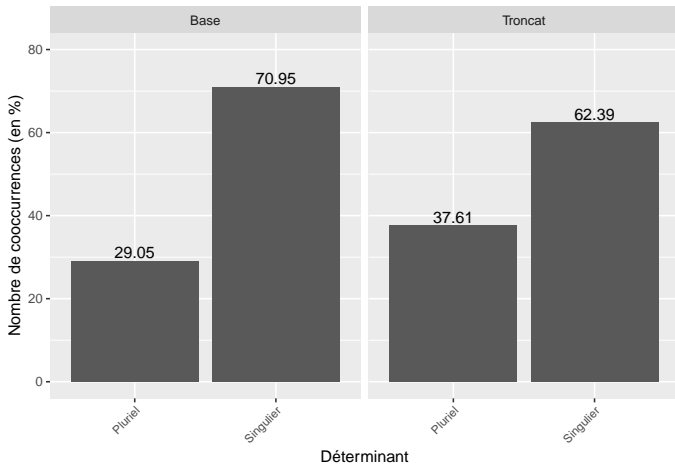
Résultats



Résultats



Résultats



Résultats

