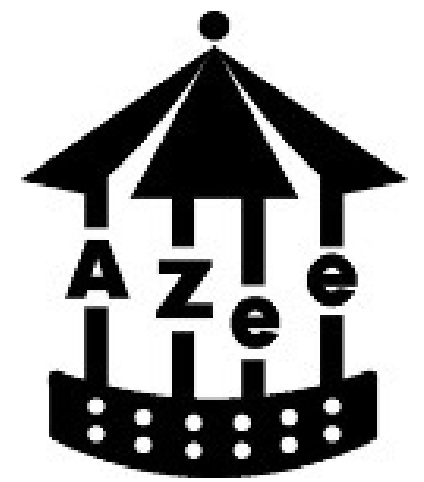


JOURNÉE D'ÉTUDES "LANGUE ET INTERFACE"
NANCY - 31/03/2023



Apports de l'informatique à la linguistique : une nouvelle grammaire pour les langues des signes ?

Camille Challant

université
PARIS-SACLAY



L1SN
LABORATOIRE INTERDISCIPLINAIRE
DES SCIENCES DU NUMÉRIQUE

LANGUES DES SIGNES

Des langues naturelles, visuo-gestuelles, multilinguistiques, mettant en jeu de nombreux articulateurs



© 2012 LIMSI-CNRS



vidéo

INTRODUCTION

VIDEO 1R-JP

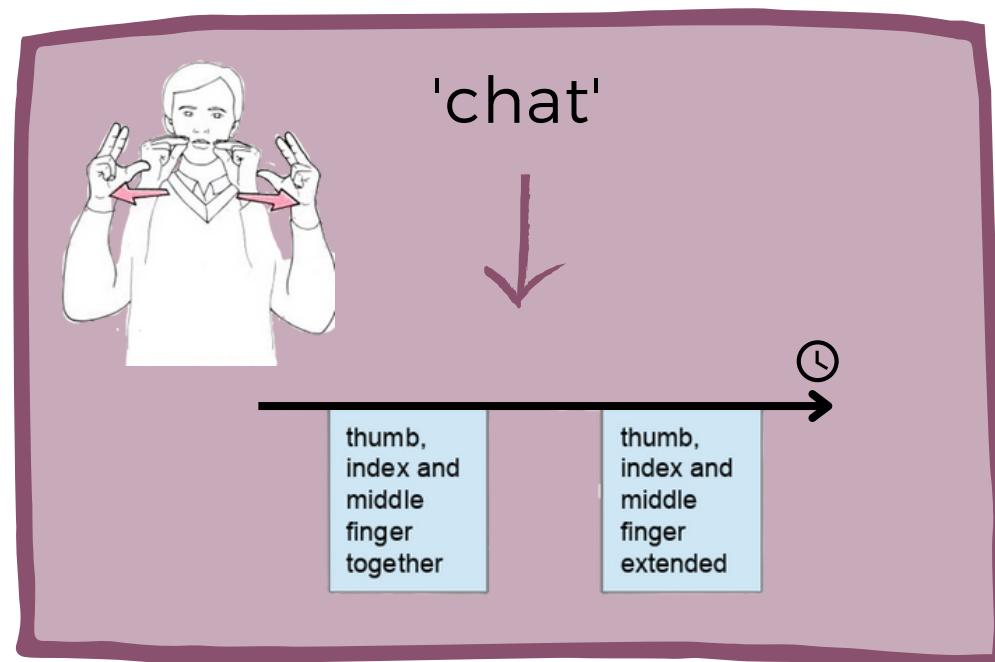
- Modélisation des langues des signes : essentielle pour différentes tâches de traitement automatique
- AZee : modèle développé au laboratoire, qui fait état de l'art en synthèse par avatar
- AZee a donc une utilité en informatique à la base, mais on découvre une portée intéressante au niveau linguistique



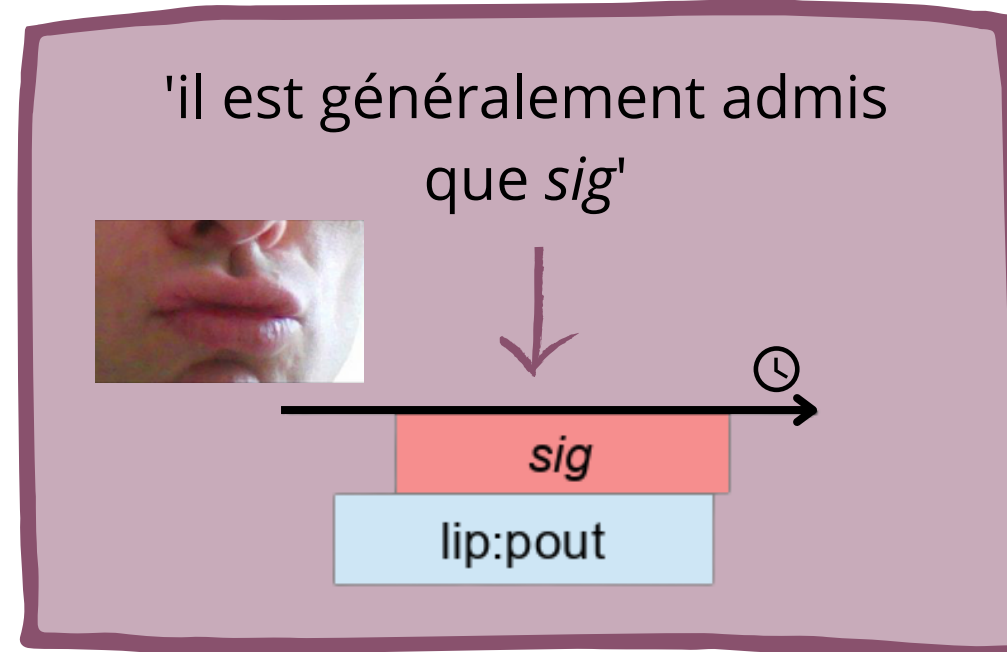
[1] : avatar Paula, université DePaul à Chicago

AZee

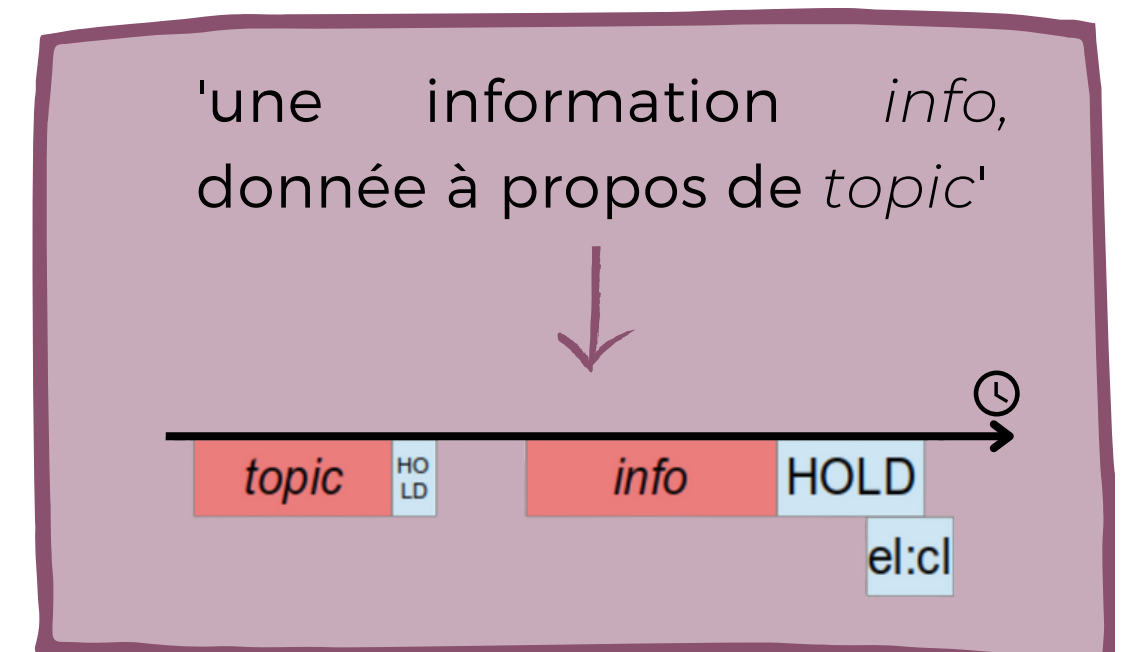
- Modèle formel basé sur la notion de **règle de production** : fonction qui représente un sens interprétable et qui détermine des formes à produire



chat



inter-subjectivity



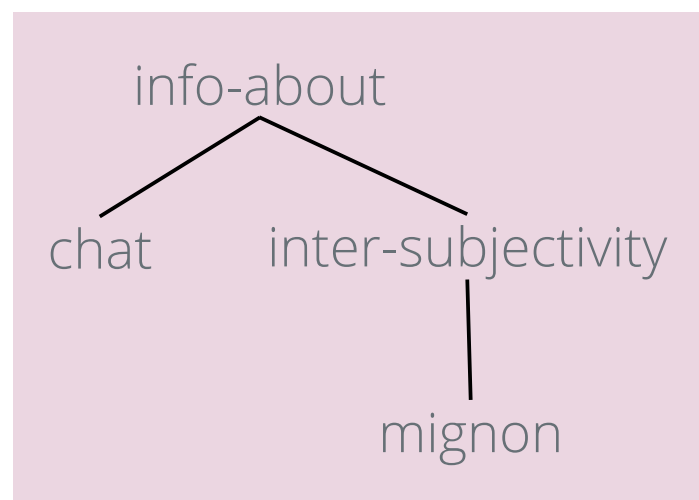
info-about

- On ne présuppose pas l'existence de niveaux linguistiques, d'une séquence ou encore de catégories syntaxiques.

AZee

Expressions AZee de discours

Combinaison de règles de production qui représente la sémantique d'un énoncé en langue des signes et qui en génère les formes



:info-about

'topic

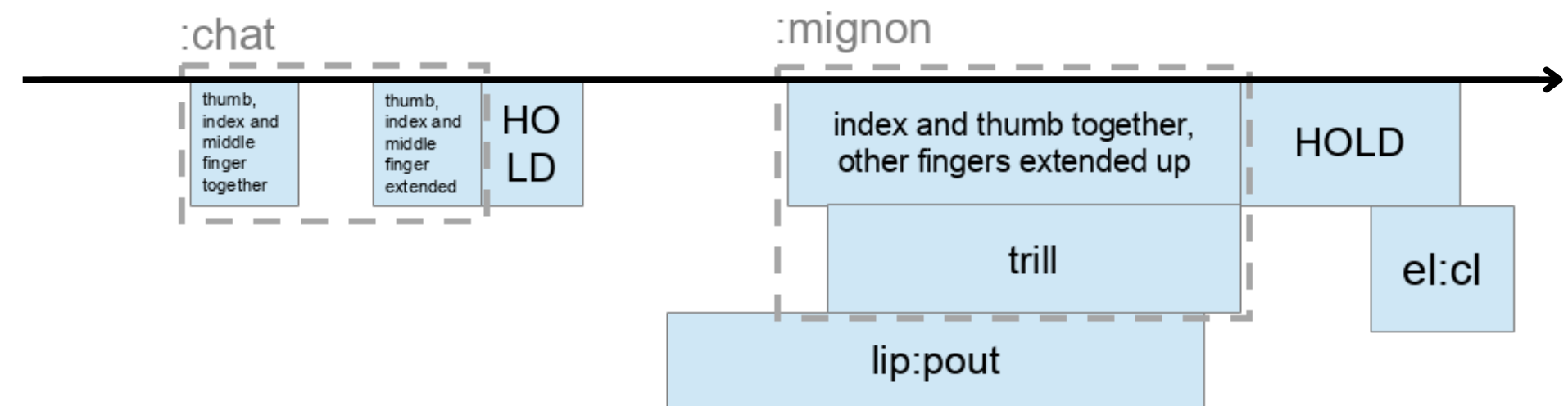
:chat

'info

:inter-subjectivity

'sig

:mignon



Les chats sont généralement admis comme mignons



Est-il possible de représenter des discours entiers en LSF avec AZee ?

SÉLECTION DU CORPUS



Corpus des 40 brèves

- 40 brèves traduites en LSF par 3 traducteurs Sourds professionnels = 120 discours, 1h de LSF
- Genre journalistique intéressant
- Bonne qualité de vidéo, montage 2 vues

TEST d'AZee SUR CORPUS

AZeefication du corpus

- 120 expressions AZee de discours
- Plus de 94% de couverture
- Corpus publié, disponible sur Ortolang

Le Français Bernard Planche, 52 ans, a été retenu en otage en Irak pendant 35 jours

1	:context	47	:pointage index
2	'ctxt	48	'target
3	:about-point	49	^Lssp
4	'pt	50	'info
5	^Rssp	51	:info-about
6	'locsig	52	'topic
7	:category	53	:âge
8	'cat	54	'info
9	:pays	55	:cpt-années
10	'elt	56	'quantity
11	:Irak	57	:tens-units
12	'proc	58	'tens
13	:about-point	59	.nb-5
14	'pt	60	'units
15	^Lssp	61	.nb-2
16	'locsig	62	'info
17	:info-about	63	:side-info
18	'topic	64	'focus
19	:side-info	65	:otage
20	'focus	66	'loc
21	:side-info	67	^Rssp
22	'focus	68	'info
23	:info-about	69	:side-info
24	'topic	70	'focus
25	:une personne	71	:temps qui passe
26	'info	72	'info
27	:France	73	:info-about
28	'info	74	'topic
29	:category	75	:cpt-unité
30	'cat	76	'num
31	:nom	77	:tens-units
32	'elt	78	'tens
33	:fingerspelling	79	.nb-3
34	'letters	80	'units
35	list	81	.nb-5
36	.B	82	'elt
37	.P	83	:jour 1
38	.L	84	'info
39	.A	85	:passer
40	.N	86	
41	.C		
42	.H		
43	.E		
44	'info		
45	:info-about		
46	'topic		

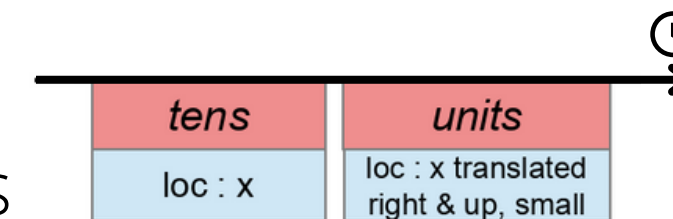
QUESTION DE RECHERCHE

Existe-t-il des contraintes qui régissent les expressions AZee de discours ?

Exemple de raisonnement :

À partir des règles de production suivantes :

- *nb-1*, qui produit la forme 👍
- *nb-2*, qui produit la forme 👉
- *tens-units(tens, units)*, qui produit des nombres à deux chiffres



4 combinaisons théoriquement possibles :

11	12	21	22
:tens-units	:tens-units	:tens-units	:tens-units
'tens	'tens	'tens	'tens
.nb-1	.nb-1	.nb-2	.nb-2
'units	'units	'units	'units
.nb-1	.nb-2	.nb-1	.nb-2

Or, dans les corpus, on observe **21** et **22**, mais jamais **11** et **12**.

On peut donc formuler l'hypothèse suivante : *nb-1* n'apparaît jamais comme premier argument de *tens-units*

ÉTUDE STATISTIQUE

Différents tests sont possibles sur une expression AZee (E) :

- $rootname(E)$ - nom de la règle de production utilisée en racine en E
- $prodcount(E)$ - nombre d'applications de règles de production dans E
- $contains(E,R)$ - vrai ssi il existe une sous-expression E' dans E telle que $rootname(E') = R$ pour R étant une règle de production

:info-about
'topic
:chat
'info
:inter-subjectivity
'sig
:mignon

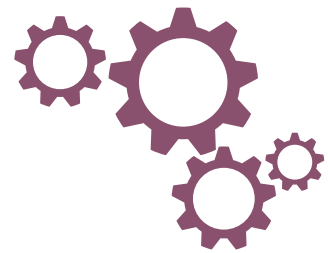


Utilisation du corpus d'expressions AZee de discours pour faire de premiers tests
(11 470 applications de règles de production)

MÉTHODE



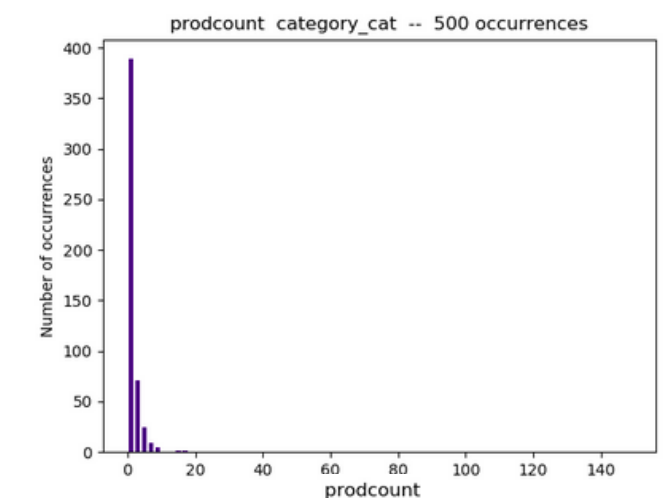
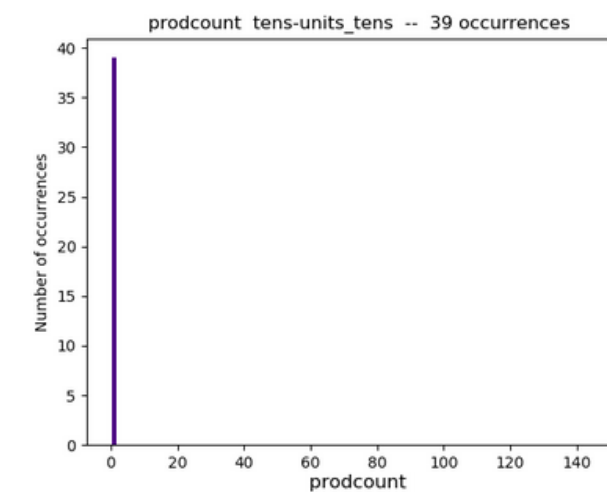
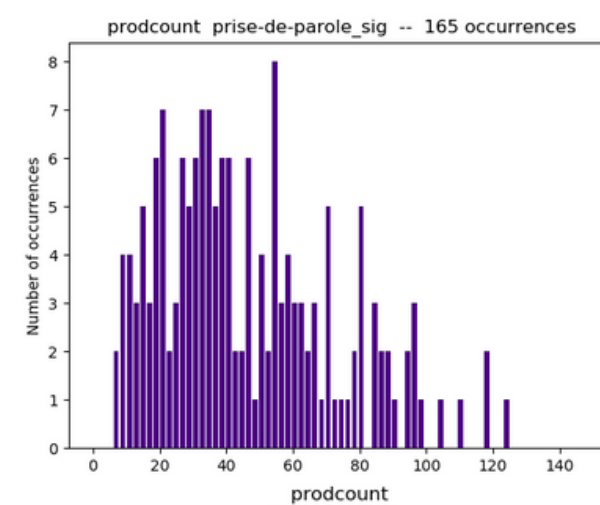
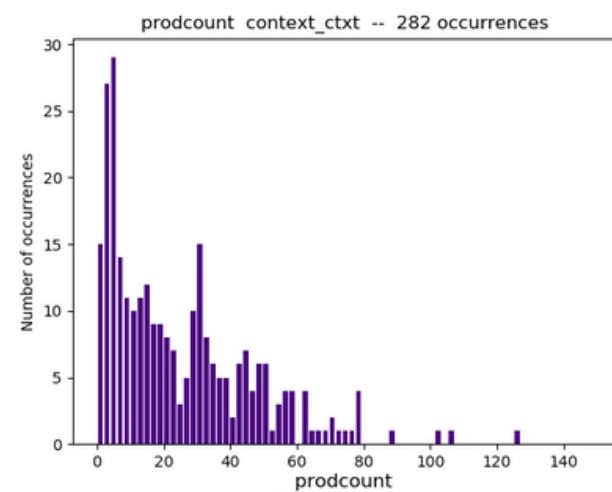
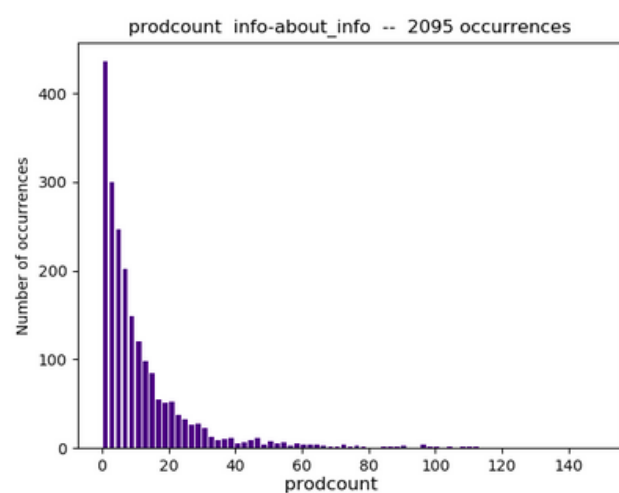
Les arguments des règles de production ont-ils tous la même complexité ?



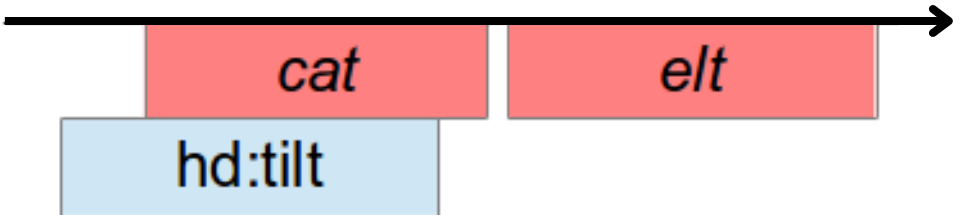
Combinaison de tests :

- *rootname*, pour trouver les expressions ayant pour racine *règle_arg*
- *prodcoun*, pour refléter la complexité de ces expressions

Résultat



category(cat, elt)

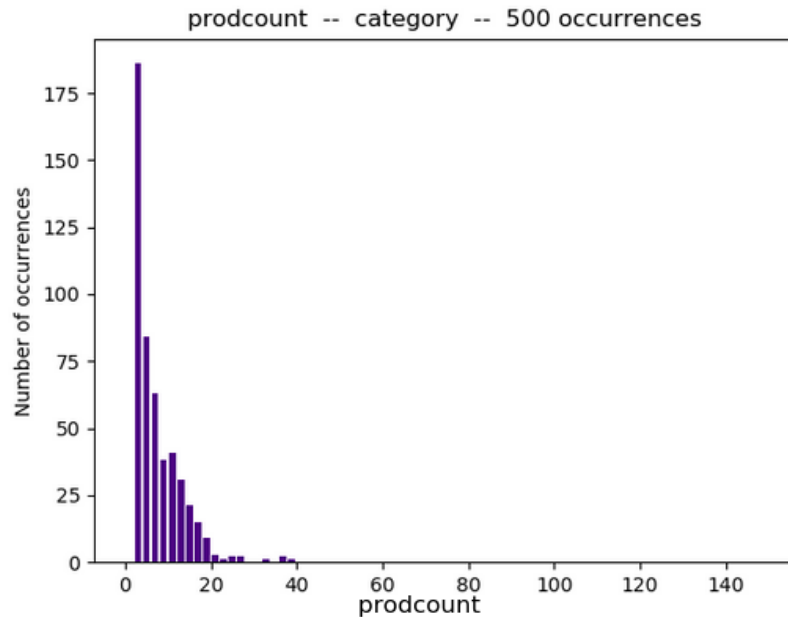


sens : *elt*, à interpréter comme une instance de *cat*
 500 occurrences dans le corpus

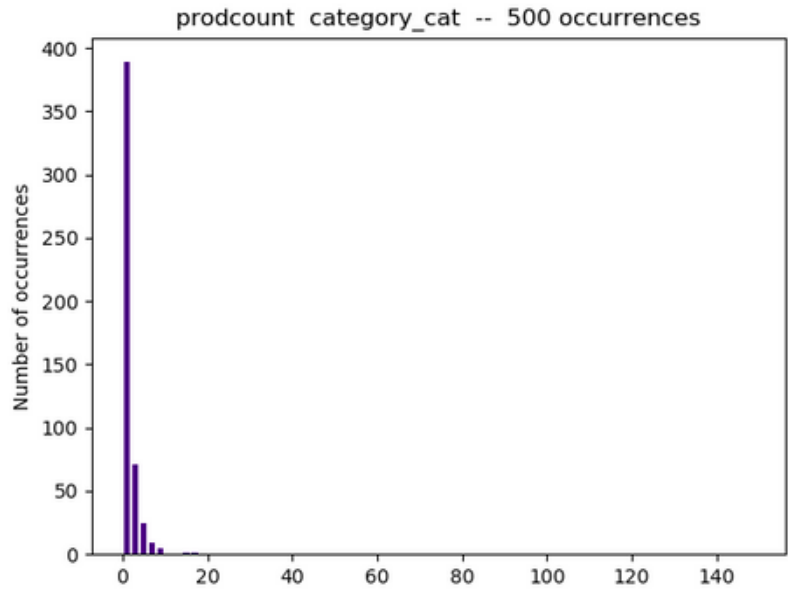
avril
 :category
 'cat
 :mois
 'elt
 :poisson

Observations :

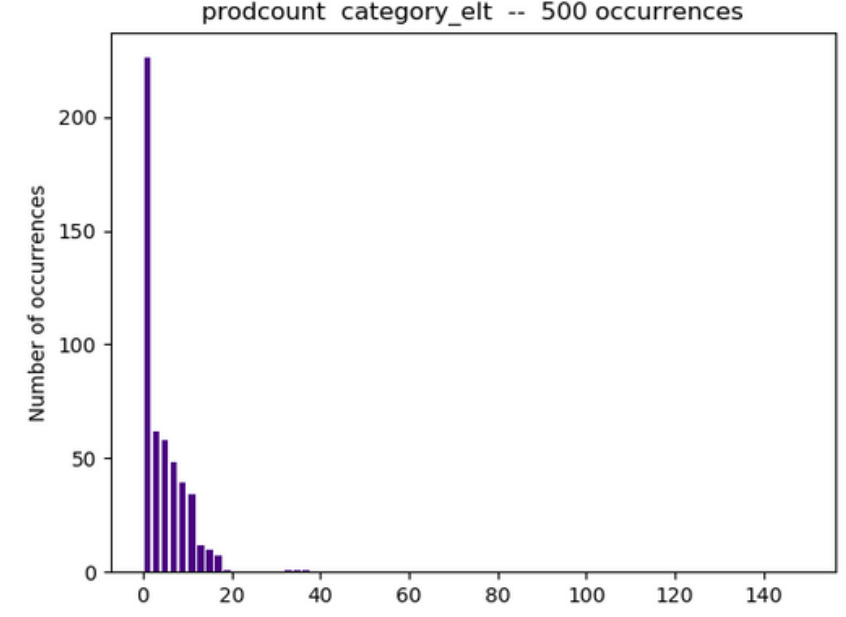
- *prodcunt*



category



cat



elt

- règles de production qui ne sont **jamais** contenues dans...

category_cat :

- context
- prise-de-parole
- nicht-sondern
- intensity
- nb-sum
- tens-units
- fingerspelling
- seq-res
- each-of

category_elt :

- context
- prise-de-parole
- nicht-sondern

prise-de-parole(sig)

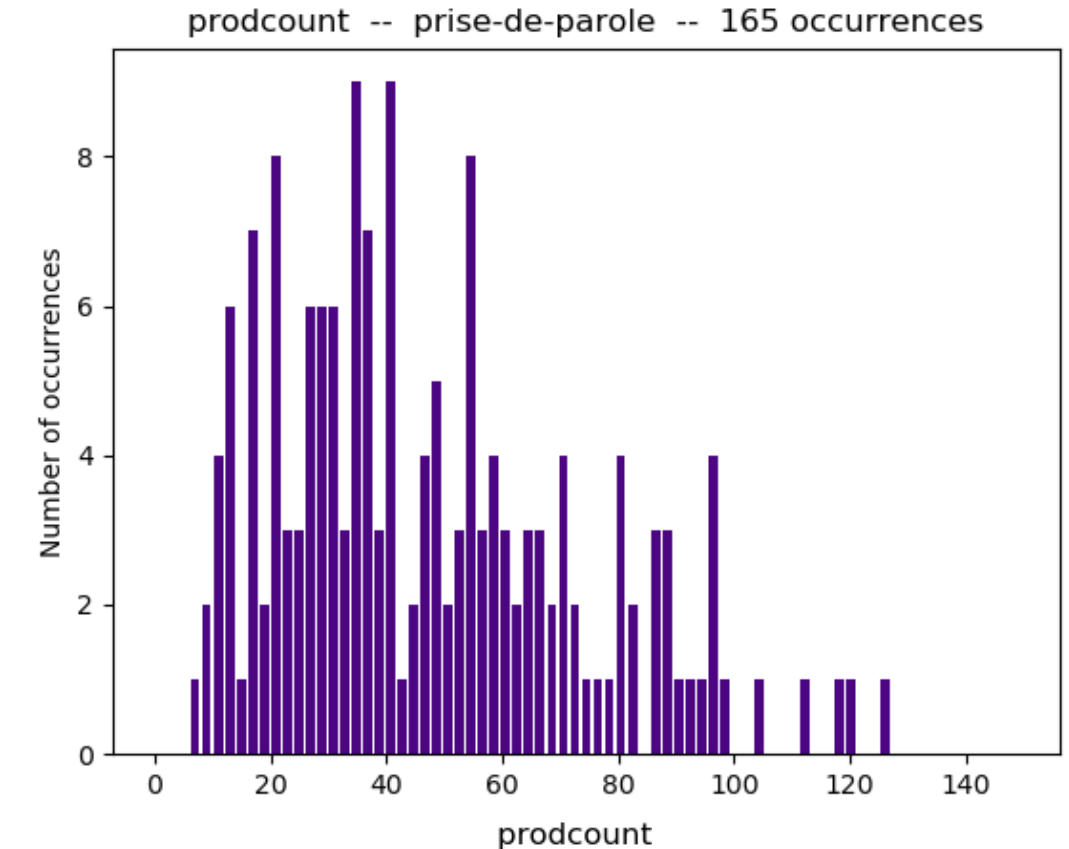
sens : pause rhétorique avant *sig*

165 occurrences dans le corpus

vidéo

Observations :

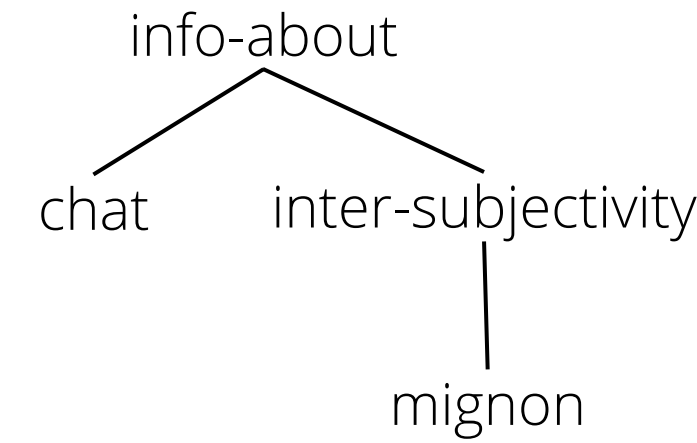
- *prodcoun*t compris entre 7 et 126
- peut contenir n'importe quelle règle de production
- n'apparaît que dans certaines positions :
 - *context_proc* : 138 occ.
 - *info-about_info* : 19 occ
 - *context_ctxt* : 4 occ.
 - *each-of_items* : 3 occ.



DISCUSSION

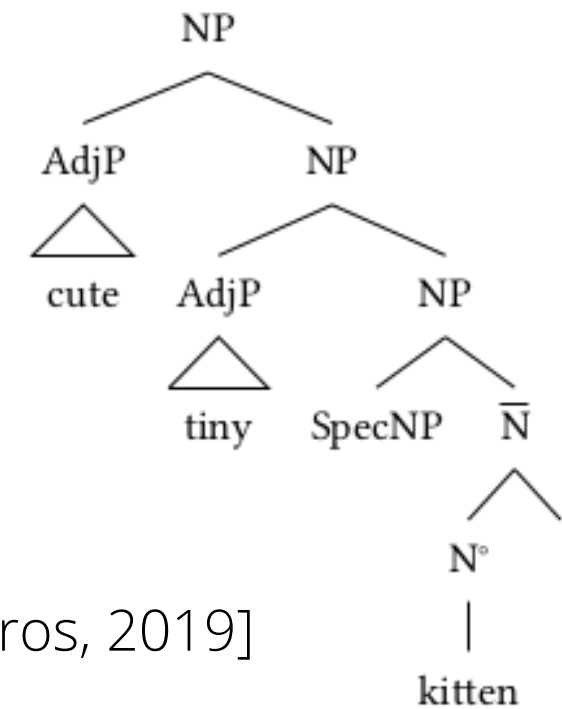
- D'après nos observations, nous pouvons émettre quelques hypothèses :
 - *category* a une limite **maximale** concernant le nombre de règles de production appliquées dans l'expression dont elle est la racine, alors que *prise-de-parole* a une limite **minimale**
 - *category* peut apparaître dans n'importe quelle position, mais pas *prise-de-parole*
 - certaines règles ne sont jamais contenues dans *category*, ce qui n'est pas le cas pour *prise-de-parole*
- Comment interpréter ces contraintes ?
 - expression AZee = représentation de la LSF \rightarrow contrainte sur l'expression = contrainte grammaticale
 - système de contraintes grammaticales = grammaire formelle

POSITIONNEMENT THÉORIQUE



- Grammaire AZee
 - Absence de plusieurs présupposés : séquence, niveaux linguistiques, catégories syntaxiques
 - Sémantique intégrée à tous les niveaux de l'expression
 - Chaque expression détermine des formes à produire (non-manuel inclus)

- Divergence avec les grammaires traditionnelles :
 - Toutes supposent des catégories syntaxiques
 - Grammaires génératives : la syntaxe est un niveau distinct de la sémantique



[Bros, 2019]

- Points communs avec :
 - les grammaires cognitives : prise en compte de la sémantique
 - les grammaires de construction : continuum lexique-syntaxe, associations forme-sens
 - les grammaires de propriétés : système de contraintes, continuum de grammaticalité

PERSPECTIVES

- Extraire automatiquement des motifs réguliers du corpus (collaboration Modyco)
- Appliquer la même méthodologie sur un corpus d'un autre genre, contenant de nombreuses structures iconiques (Mocap 1)
- Confirmer les conjectures : générer avec un avatar des énoncés qui ne répondent pas à nos contraintes et les présenter à des locuteurs natifs

