

CHUNKING GUIDELINES V2

1. Chunks

Definition : The chunk is the smallest linguistic's item string possible, and a non recursif component. It include a head and other items that depends of it. The division in chunks allow a first segmentation that show some potential macro-syntactic borders. This type of annotation is particularly usefull for spoken language. As we know spoken language can be hard to analyse syntactically, specially because of those non ended sentences.

UNKNOWN : chunk de nature non-identifiée (amorce, mot inconnu)
AP : adjectival phrases (adjectif après le nom)
AdP : adverbial phrases
CONJ : conjunction
NP : nominal phrases (tête = nom ou pronom non clitique : toi, moi)
PP : prepositional phrases
VP : verbal phrases
SENT : punctuation
ARTIC : articulateur
FNO

2. Liste de figement

1.2. Adverbes/ARTIC composés

à fond
à la fin
à peu près
au moins
au pire des cas
bien entendu
d'autrefois
de nos jours
de suite AdP
de toute façon
du tout
d'habitude
en effet
en fait
en fin de compte
en plus
en tout cas
entre autres
là dedans
malgré tout
mine de rien
non plus
par exemple
petit à petit
plus ou moins
pour le coup
quand même
tant pis
tout à fait
tout simplement
un petit peu
un peu
un tout petit peu

2.2. ARTIC composées

ainsi que
alors qu'
au cas où
du coup
d'autant plus
encore que
et puis
même si
ou bien
parce que
pour que

sauf que CONJ
vu qu'
comment dire
du coup
eh ben
entre guillemets
et bien
n'empêche
n'est ce pas

2.3. Prépositions composées

à la fin de
à l'intérieur du
à l'occasion de
à partir de
à propos de
au coeur de
au cours des...
au début du
au dessus de
au moment des
au niveau des
aux travers des
en plus des
en train de
jusqu'à
loin de
mis à part [PP 0] le [PP 1] terroir [PP 2]
par l'intermédiaire des

2.4. Noms composés

bateau à vapeur
Amérique du nord
quelque chose [NP 0]
gestionnaire de paies [NP 0] => métier
qu'est-ce qui [NP 0]
la plupart d' [NP 0] entre [NP 1] eux [NP 2]
vins médecin

2.5. Déterminants composés

dix huit cent quatre-vingt-dix
plusieurs(déterminant quantitatif) + ... + de suite (figé) :
pendant [PP 0] plusieurs [PP 1] semaines [PP 2] de suite [PP 3]

3. Articulateur

Articulateur réunit plusieurs catégories : MD, IP, relatif, conjonctions, certains adverbes

C'est une catégorie fonctionnelle, non autonome (vs FNO). Ce sont des éléments qui organisent le contenu propositionnel, permettent la cohésion aux différents niveaux :

a). interactionnel, entre différents tours de parole (MD, IP)

b). à l'intérieur d'un tour de parole

- textuel : il est grand **mais** il est bon, il est grand **en effet** il est bon

- syntaxique : je dis **que** tu vieillis bien.

Il n'est pas autonome que ce soit syntaxiquement ou interactionnellement, contrairement à FNO.

La présence de multiples articulateurs montre que l'oral articule beaucoup.

Souvent il est suivi ou précédé d'un FNO ce qui s'explique par son non autonomie et par l'autonomie d'un FNO. Parmi les adverbes articulateurs, on n'aura jamais les adverbes modifieurs d'un syntagme (très, par ex.).

Les exemples des articulateurs : mais, du coup, donc, enfin, alors, style, genre, c'est-à-dire, en effet, en fait, par exemple, n'est-ce pas, ou bien, que, dont, où, entre guillemets, et puis, comment dire, en tout cas, entre autres, troisièmement, bien entendu, mine de rien, malgré tout, en gros, encore que etc.

(Modalisateur : position du locuteur par rapport à ce qu'il dit)

Articulateur peut avoir un modifieur (tout comme adjectif, nom, etc.) :

hé [ARTIC 0]

bien [ARTIC 0]

tout simplement [ARTIC 0] parce que [ARTIC 1]

4. FNO

FNO (forme noyau) : oui, ouais, ouf, merde, d'accord, voilà, bonjour, salut, carrément, mince, allez, santé, merci, purée, putain, non, hélas, bref, pardon

5. Ambiguïté

1.2. D'accord

d'accord : AdP ou FNO

je suis d'accord => AdP

Tu viens déjeuner ? D'accord => FNO

L'adverbe n'est pas toujours une tête : très beau

très/fort/bien/un peu/trop + Adv => AdP (un seul chunk)

très/fort/bien/un peu/trop + Adj => AP

5.2. Participe passé / adjectif

les [NP 0] sarments [NP 1]

étaient [VP 0]

raccrochés [AP 0]

5.3. Préposition / adverbe

5.3.1. après

ELI elle [VP 0] est [VP 1]

ELI trop [AP 0] crevée [AP 1]

ELI après [AdP 0]

ELI le [NP 0] vendredi [NP 1] soir [NP 2]

=> il y a une pause, hésitation, écoute a choisi

et une préposition dans d'autres cas.

5.3.2. au début

ELI parce qu' [ARTIC 0]

ELI au début [AdP 0]

ELI tu [VP 0] sais [VP 1]

au début de = prep

5.3.3. à la fin de

ELI de [PP 0] t' [PP 1] embaucher [PP 2]

ELI à la fin [AdP 0]

-

ELI normalement [AdP 0]

ELI à la fin de [PP 0] l' [PP 1] année [PP 2]

5.4. FNO/ADV/ARTIC

AdP/ARTIC

ELI ça [VP 0] se [VP 1] passait [VP 2]

ELI bien [AdP 0]

ELI et [ARTIC 0]

ELI tout [NP 0]

ELI mais [ARTIC 0]

ELI **déjà** [ARTIC 0]
ELI elle [VP 0] a [VP 1] mal [VP 2] compris [VP 3]
ELI ce que [NP 0]
ELI c' [VP 0] était [VP 1]

ELI je [VP 0] croyais [VP 1]
ELI **aussi** [AdP 0]
ELI que [ARTIC 0]
ELI c' [VP 0] était [VP 1]
ELI faire [VP 0]
ELI des [NP 0] paies [NP 1]

/

c' [VP 0] est [VP 1]
aussi [ARTIC 0]
une [NP 0] interprétation [NP 1] un peu [NP 2] personnelle [NP 3]

FNO/ARTIC

oui, non, si

BEA **non** [ARTIC 0]
BEA ouais [FNO 0]

FNO/AdP

ELI je [VP 0] vous [VP 1] sers [VP 2]
MAR carrément [FNO 0]
vs

ELI je [VP 0] suis [VP 1]
ELI pas [AP 0] sûre [AP 1]
ELI carrément [AdP 0] plus [AdP 1]

PRO/MD : quoi,

ADJ/MD : bon

6. Précisions

1.2. Répétition

Répétition (des clitiques) : mettre systématiquement la répétition (des préposition et des déterminants) à l'intérieur d'un chunk (peu importe si c'est la tête ou pas)

de [PP 0] de [PP 1] survivre [PP 2]

ces [NP 0] ces [NP 1] petites [NP 2] bêtes [NP 3]

à [PP 0] le [PP 1] à [PP 2] le [PP 3] boire [PP 4]

de [PP 0] tout [PP 1] euh [PP 2] tout [PP 3] de [PP 4] de [PP 5] tout [PP 6] montrer [PP 7]
se [VP 0] s' [VP 1] ajouter [VP 2]
on [VP 0] se [VP 1] on [VP 2] constate [VP 3]

6.2. décomposer (pas figé)

6.2.1. **Avoir+N**

avoir + N : on a considéré que ce n'était pas figé car o

a [VP 0]

tendance [NP 0]

on a considéré que ce n'était pas complètement figé car on peut dire : a une grande tendance, je vois chez elle une tendance (sans le verbe avoir).

6.2.2. **A ce moment-là**

à [PP 0] ce [PP 1] moment-là [PP 2]

pas complètement figé non plus car on peut dire à cette étape, etc. une structure productive (critère)

d' [PP 0] une [PP 1] manière [PP 2]

6.3. Relatifs

Font partie des articulateurs

qui, où, dont [ARTIC 0]

ce [NP 0]

qu' [ARTIC 0]

6.4. Disfluences

interjections, MD font partie des articulateurs mais peuvent être inséré dans un autre chunk, dans ce cas ils ne font plus la tête d'un chunk

Deux cas à distinguer

a). un chunk autonome : ARTIC

b). au milieu des chunks

avec [PP 0] des [PP 1] cargaisons [PP 2] euh [PP 3] complètes [AP 4]
des [NP 0] comment dire [NP 1] des [NP 2] camions [NP 3] entiers [NP 4]
une [NP 0] euh [NP 1] comment dire [NP 2] un [NP 3] rapprochement [NP 4]
de [PP 0] euh [PP 1] de [PP 2] l' [PP 3] espace [PP 4] viticole [PP 5]
de [PP 0] leur [PP 1] de [PP 2] ouais [PP 3] leur [PP 4] journée [PP 5]

j' [VP 0] ai [VP 1] euh [VP 2] en quelque sorte [VP 3] été [VP 4] intronisé [VP 5]

6.5. Adverbes

6.5.1. **Adv + adv**

Deux cas à distinguer :

a). pas tellement, absolument partout => ADV 0+ ADV 1

pratiquement [AdP 0] pas [AdP 1]

=> une tête

b). ici [AdP 0]

véritablement [AdP 0]

=> deux têtes

6.5.2. **Trop**

Il est modifieur

trop+adj/adv/nom=>chunk adj/adv/n

trop [NP 0] la [NP 1] merde [NP 2]

6.6. Entités nommées

des [PP 0] Appellations d'Origine Contrôlée [PP 1]

Centre Val de Loire [NP 0]

RGA

6.7. UNKNOWN

- qui [NP 0]

le [_UNKNOWN_]

qui [NP 0]

l' [VP 0] ont [VP 1]

- plus ou moins [AdP 0]

de [PP 0] vins médecin [PP 1]

le [_UNKNOWN_]

euh [IP 0]

certaines [NP 0] mauvaises [NP 1] années [NP 2]

le clitique ne peut pas être une tête de chunk.

6.8. Chiffres

de [PP 0] deux mille [PP 1]

mille neuf cent cinquante-huit [NP 1]

6.9. Infinitif

ELI je [VP 0] veux [VP 1]

ELI dire [VP 0]

ELI tu [VP 0] vas [VP 1]

ELI avoir [VP 0]

On sépare quand même, même si d'un côté on a une construction lexicalisée (vouloir dire) de l'autre grammaticalisé/

vs

je regarde passer les enfants

MAR on [VP 0] nous [VP 1] entend [VP 2]

MAR arriver [VP 0]

6.10. Conjonction

La conjonction fait partie d'un chunk articulateur sauf au cas où elle coordonne les adjectifs à l'intérieur d'un chunk nominal. Plusieurs cas à distinguer :

- (*et*)_{ARTIC} (*toi*)_{NP}

- (*une belle et jolie fille*)_{NP}

- (*une fille*)_{NP} (*belle*)_{AP} (*et*)_{ARTIC} (*intelligente*)_{AP}

- (*une belle*)_{NP} (*et*)_{ARTIC} (*une jolie fille*)_{NP}

6.11. Participe passé

Verbe au participe passé ne peut pas être une tête d'un chunk et fait partie toujours d'un chunk verbal de son auxiliaire qui le précède. Les adverbes de négation ou autres qui s'insèrent entre l'auxiliaire et le participe feront partie d'un chunk verbal : (vous n'avez pas vu)_{VP}, (vous avez vu)_{VP}, (vous avez très bien vu)_{VP}

7. Règles de concaténation

- VP+VP(inf)=>VP

je [VP 0] vais [VP 1]

passer [VP 0]

- VP+PP=>VP

- NP+PP=>NP

- PP+PP=>PP ???

par [PP 0] aussi [PP 1] ce [PP 2] ce [PP 3] biais [PP 4]

de [PP 0] la [PP 1] religion [PP 2] chrétienne [PP 3]

sur [PP 0] euh [PP 1] la [PP 2] géographie [PP 3]

de [PP 0] la [PP 1] France [PP 2]

-plusieurs ARTIC 0 : ensemble

et [CONJ 0]

ELI donc [ARTIC 0]

ELI du coup [ARTIC 0]

ELI donc [ARTIC 0]

ELI oh [ARTIC 0]

ELI ben [ARTIC 0]

/

hé [ARTIC 0]

bien [ARTIC 0]

tout simplement [ARTIC 0] parce que [ARTIC 1]

=> on ne regroupe que hé+bien