

Liste des propriétés de mots disponibles dans les corpus de la Base de français médiéval

Source : Annexe 1 du *Tutoriel de la BFM*

Le tableau ci-dessous donne la liste des propriétés de mots disponibles et utiles pour l'interrogation des différents corpus de la BFM. La colonne « Corpus » indique les corpus dans lesquels les propriétés sont disponibles :

- BFM2019 : corpus BFM2019, PALAFRAFRO-V2-2 et PALAFRAPAR
 - BFM2016 : BFM2016 et CORPTEF (ces propriétés disparaîtront dans les nouvelles versions des corpus)
- tous : tous les corpus français

| Nom de la propriété | Explication | Exemple de requête | Corpus |
|---------------------|--|--|----------------------|
| cattex-pos | Étiquette morphologique Cattex 09 | [cattex-pos="ADJ.*"] | PALAFRA |
| cattex-pos_syn | Étiquette morphosyntaxique Cattex 09, identique à cattex-pos, sauf dans les cas de substantivation, etc., identifiés dans certains textes vérifiés | [cattex-pos_syn="NOMcom" & cattex-pos="VER.*"] | PALAFRA |
| etiquetage | Indique si l'étiquetage morphosyntaxique a été vérifié pour une occurrence donnée | [etiquetage="vérifié"] | PALAFRA |
| frolemma | Identique à la forme (peu utile pour les requêtes) | | BFM2016 |
| fropos | Étiquette morphosyntaxique automatique non vérifiée (Cattex 09) | [fropos="ADJ.*"] | BFM2016 ¹ |
| id | Identifiant du mot dans le corpus (peu utile pour les requêtes) | | tous |
| lang | Langue du mot étranger | [lang="la.*"] | tous |
| lemma | Lemme de l'occurrence | [lemma contains "avoir"] | BFM2019 |
| lemma_src | Source du lemme, voir la section 3 | [lemma_src="BFM"] | BFM2019 |
| lemmatisation | Indique si le lemme est vérifié ou non | [lemmatisation="vérifié"] | BFM2019 |
| orig | Indique si le mot est balisé <orig> (graphie non régularisée) dans le document XML-TEI, utilisée pour certaines ponctuations dans la <i>Queste del saint Graal</i> | [orig!="NA"] | tous |
| pos | Étiquette morphosyntaxique Cattex 09 | [pos="ADJ.*"] | BFM2016 |

¹ La propriété `fropos` est également disponible dans le corpus BFM2019 pour le confort des utilisateurs, sa valeur est identique à `cattex-pos` (étiquette Cattex 09 vérifiée ou non vérifiée).

| | | | |
|----------|--|--|---------|
| | vérifiée | | |
| q | Niveau d'imbrication de l'occurrence dans un discours direct (« 0 » si hors du discours direct, « 1 » si dans un discours direct non imbriqué, « 2 » ou « 3 » si dans un discours direct imbriqué dans un ou deux autres discours directs) | [q="[123]"] | tous |
| ref | Référence par défaut pour la concordance (peu utile pour les requêtes) | | tous |
| sic | Indique si le mot est balisé <sic> (graphie visiblement erronée) dans le document XML-TEI, | [sic!="NA"] | tous |
| sp | Indique si le mot se trouve dans une prise de parole d'un texte dramatique ou d'un dialogue (balise TEI <sp>), valeurs possibles : « t » (vrai) ou « f » (faux) | [sp="t"] ou [sp="f"] | tous |
| supplied | Indique si le mot se trouve dans un passage balisé <supplied> (conjecture éditoriale) dans le document XML-TEI, valeurs possibles : « t » (vrai) ou « f » (faux) | [supplied="f"] | tous |
| ud-feats | Tag for all UD features | [ud-feats contains "PronType=Det"] | BFM2019 |
| word | La forme du mot, propriété par défaut. Une syntaxe raccourcie peut être utilisée si elle n'est pas combinée avec d'autres propriétés | a, "a" ou [word="a"] [word="a" & cattex-pos="VERCjg"] | tous |