

# MWEs em UD

19/04/2022  
Elvis de Souza

# Diretivas UD

<https://universaldependencies.org/v2/mwe.html>

UD não permite “palavras com espaço”. Por isso, propõem aumentar a consistência usando três classes:

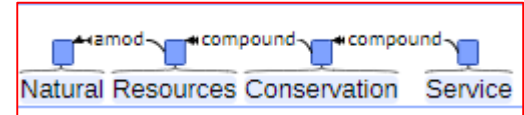
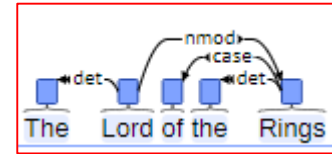
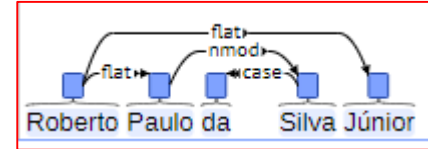
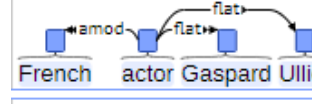
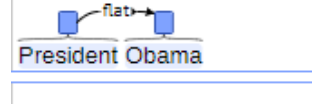
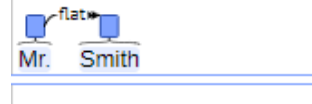
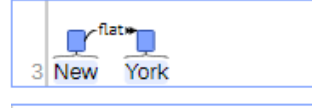
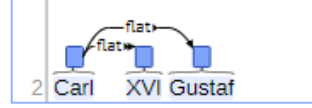
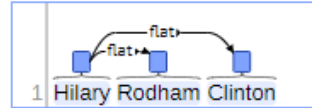
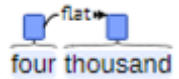
- fixed (expressões completamente fixas)
- flat (expressões sem governante sintático claro) e seus subtipos (*flat:foreign* e *flat:name*)
- compound (expressões com governante sintático) e seus subtipos (*compound:prt* e *compound:lvc*)

# nomes, datas e números complexos

flat: se não tiver uma estrutura sintática regular (não tem governante/head)

compound: tem governante, mas é estrutura de “dois nominais”

transparente: tem governante e não é estrutura de “dois nominais” (tem palavras funcionais)



- Dúvidas:
  - Arraial do Cabo
  - Cabo Frio
  - São Paulo
  - Roberto Júnior
  - Henrique V

# compound e a estrutura “dois nominais”

[https://universaldependencies.org/workgroups/newdoc/two\\_nominals.html](https://universaldependencies.org/workgroups/newdoc/two_nominals.html)

- “phone book”: nmod ou compound? Se houver critério de distinção, ok. Caso contrário, nmod.
- “president Obama”: flat ou appos? **flat:name** (appos deve poder ser invertido)
- “Obama, the president of the US”: appos

No PetroGold:

- uma lista muito pequena de elementos como *compound* (“óleo diesel, matéria prima”)
- “Presidente Obama”: flat:name ✓
- “presidente Obama”: nmod
- “presidente da América”: nmod ✓
- “The Lord of The Rings”: flat:name

\* Verificar se é possível colocar as palavras funcionais dependentes do token à direita (mas mantendo flat do token à direita para o da esquerda)

# compound:prt e compound:lvc

**compound:prt** não se confunde com o verbo pronominal (expl:pv – “encontra-se” ou “localiza-se” em certos contextos)

- for particle verbs (with the subtype `compound:prt`):



## compound:lvc

<https://universaldependencies.org/u/dep/compound-lvc.html>

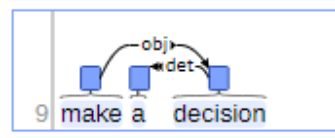
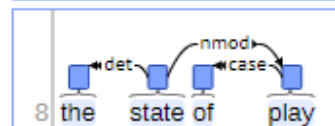
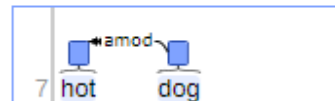
(UD): Em construções de verbo leve, o verbo não tem muito conteúdo semântico. O sentido da construção está na palavra que não é governante, geralmente um substantivo ou adjetivo.

Não há exemplos em inglês nas diretivas. Em português, seriam expressões como:

- dar um grito
- ter em mente
- tirar um cochilo
- tomar uma decisão
- fazer vista grossa

(UD): Línguas que usam a classe *compound* devem desenvolver critérios morfosintáticos (e não lexicais ou semânticos) para identificar esses termos. Por isso, as expressões à direita “parecem ser melhor descritas” de forma transparente.

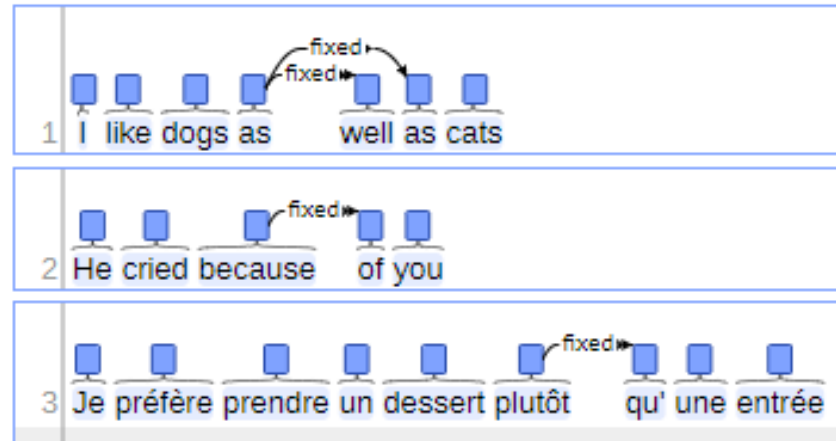
<https://universaldependencies.org/u/dep/compound.html>



# fixed: exclui construções semi-fixas e flexíveis

Funcionam como palavras funcionais ou  
adverbiais curtos

POS das palavras isoladas  
REL da expressão fixa



# lista de expressões fixas

- POeTiSA
  - lista de “non-ambiguous co-occurring tokens”
  - entradas sugeridas por linguista dedicado à anotação de corpus em UD
  - quase todas as entradas são fixed?!
- PetroGold (19 teses e dissertações gold)
  - PALAVRAS => Bosque-UD => Stanza => PetroGold
  - Gramáticas do português
  - Análise humana: ausência de composicionalidade
- Petrolês (400 documentos acad. brutos)
  - PREP (DET) NOUN PREP
  - Frequências brutas e Log-likelihood
  - Análise humana: ausência de composicionalidade

