**Adaptando a openWordnet-PT para o domínio Jurídico Brasileiro**

Daniel Lopes da Silva Ferreira Oliveira

Matheus Moro Borghi

[--]

**I - INTRODUÇÃO**

O presente relatório aborda a análise de palavras presentes em provas de primeira fase da Ordem de Advogados de Brasil (“OAB”) e as suas respectivas triagens e interpretações na Wordnet Brasil. O objetivo é encontrar erros, inexatidões, brechas ou quaisquer pontos que poderiam servir para incrementar essa biblioteca digital.

Para isso, foram recortados os textos de mais de cinco provas do exame da ordem, calculadas as palavras mais repetidas e, com base nesse dado, foi gerado um ranking de ocorrências. Em sequência, as palavras foram cruzadas com a biblioteca da Wordnet para se averiguar quais seriam as suas possíveis interpretações. O próximo passo, qual seja a análise efetiva do output obtido, será abordado com mais detalhes ao longo desse relatório.

II – Metodoogia

1. Usamos provas OAB como corpus
2. AntConc para encontrar mais frequentes. ?? Vantagens e problemas
3. Usamos freeling para análise : Por que? Freeling dá POS e LEMMA, e freeling faz WSD. Logo podemos ver rápido quais palavras já estão na OWN-PT. Problemas: Freeling erra e temos que lidar com isso.

**II – ANÁLISE**

1. Freeling acertou POS e LEMA
   1. Acertar WSD
   2. Errar WSD
2. Dada palavra P (POS, LEMA) -> OWN-PT?
   1. Palavras que tem synbset -> colocar no synset
   2. Palavras que exprimem um sentido que não tem synset na OWN-PT => criar novo synset
   3. Algum synset da PWN parece inadequado! Glossa, o link !!

A partir dos cruzamentos feitos entre as palavras e o banco de dados digital, constataram-se três possíveis resultados:

1. match total entre o significado obtido e a interpretação correta da prova (*“Match Total”*); => freeling acertou LEMA e POS. Existe na OWN-PT um synset (conceito) adequado para o significado da palavra em questão no corpus naquele contexto. E o freeling acertou a WSD.
2. interpretação da Wordnet não bate com a da prova (*“Interpretação Inexata”*); =>
3. (iii) o freeling está mapeando errado (*“Erro do Freeling”*); e (iv) a palavra não foi triada pela Wordnet (*“Triagem Inexistente”*).

Uma vez que o nosso objetivo é buscar e melhorar ineficiências, vamos focar somente nos últimos dois resultados.

1. Interpretação Inexata

* Juízo

No direito, juízo não é somente a atividade de julgar e emitir determinada opinião de valor. Juízo também pode representar “*o órgão do poder judiciário [competente] para realizar as atividades de direito*”[[1]](#footnote-1)

No entanto, conforme comentado, a definição da Wordnet mais recorrente é a seguinte:

|  |
| --- |
| [05837957-n](http://wnpt.brlcloud.com/wn/synset?id=05837957-n) mind, judgment, judgement **| ideia, opinião, juízo** |

* Recurso

No direito, Recurso *“é o instrumento utilizado pela parte vencida ou por terceiro prejudicado para provocar o reexame de uma decisão, a fim de que se promova a sua reforma, invalidação, integração ou simples esclarecimento”*[[2]](#footnote-2). Ou seja, é um instrumento, conjunto de documentos e peças.

A Wordnet possui uma definição bastante semelhante, mas imprecisa.

|  |
| --- |
| [01208597-n](http://wnpt.brlcloud.com/wn/synset?id=01208597-n) recourse, resort, refuge **| recurso** |

A sua imprecisão se deve ao fato de que a interpretação dada confere à recurso o sentido de uma ação, o ato de alguém de recorrer à uma instância superior para a reavaliação de determinada decisão. No entanto, no direito, a palavra recurso, como já mencionado, em muitas ocasiões não se refere ao ato, mas ao próprio instrumento que viabilizou o referido ato.

* Execução

No direito brasileiro, execução se refere ao “*conjunto de atos destinados à obtenção, pelo credor, em juízo, do cumprimento de obrigação constante em título executivo judicial ou extrajudicial*”[[3]](#footnote-3). Ou seja, é o procedimento judicial para viabilizar o cumprimento de determinado crédito. Caso alguém detenha um direito em título executivo, e o devedor não promova o seu pagamento, esse credor pode mover processo de execução para forçar que o devedor prossiga com sua devida obrigação.

Na Wordnet, o mapeamento mais próximo é o seguinte:

|  |
| --- |
| [00090076-n](http://wnpt.brlcloud.com/wn/synset?id=00090076-n) foreclosure **| execução hipotecária** |

Esta palavra, conforme definida na Wordnet, é um termo que estaria contido em “Execução”. Após um acesso rápido ao mapeamento conferido a ela, percebe-se que ela possui “Relações” que seriam também comuns à “Execução”, caso esta tivesse sido devidamente mapeada:

|  |
| --- |
| Relations  * Lexicographer file: (noun.act) * RDF Type: (NounSynset) * Hyponym of: [ [legal\_proceeding](http://wnpt.brlcloud.com/wn/synset?id=01184814-n) ] |

### A partir da definição previamente mencionada, “Execução” também seria um procedimento legal.

* Concessão

No direito brasileiro, “concessão” em geral é *“o acordo de vontades entre a Administração Pública e um particular, pelo qual a primeira transfere ao segundo a execução de um serviço público, para que este o exerça em seu próprio nome e por sua conta e risco, mediante tarifa paga pelo usuário”[[4]](#footnote-4).*

Assim, seria basicamente um acordo ou contrato realizado entre a Administração Pública e um particular para que este possa prestar serviços públicos, que são de titularidade daquela. Na wordnet, o mapeamento mais próximo da definição jurídica é o seguinte:

|  |
| --- |
| [06526619-n](http://wnpt.brlcloud.com/wn/synset?id=06526619-n) grant, concession **| concessão** |

No entanto, a sua definição diverge completamente da supracitada:

|  |
| --- |
| *(a contract granting the right to operate a subsidiary business; "he got the beer concession at the ball park")* |

Ao meu ver, a definição esboçada acima, ao menos em português, não está correta. “Concessão”, somente, não remete a um contrato para operar um negócio ou empresa subsidiária. Por conseguinte, nesse caso, sugere-se não a criação de um novo *synset*, mas a atualização deste *synset* mencionado para a definição correta em português.

* Nulidade

Nulidade em geral é usada, no direito, para indicar que determinado ato, por ferir determinação legal. Se a nulidade for absoluta, ela não deveria jamais ter nascido. Assim, a nulidade serve para extinguir quaisquer efeitos que esse ato tenha produzido, retroativamente ao momento de sua criação (efeitos “*ex tunc*”). É como se o ato jamais tivesse existido. No entanto, caso possa ser sanável, diz-se que a nulidade é relativa, e o ato na verdade é apenas anulável (efeitos “*ex nunc*”).

Em síntese, “nulidade” *“é a ineficácia do ato ou relação processual, causada pela não observância da lei”[[5]](#footnote-5)*.

Na Wordnet, o único mapeamento para essa palavra é o seguinte:

|  |
| --- |
| [02497471-a](http://wnpt.brlcloud.com/wn/synset?id=02497471-a) ineffectual, unavailing, futile, otiose **| nulidade, fútil, inútil, vão** |

Embora seu significado tenha certa intercessão, este *synset* não expressa a correta definição do termo jurídico.

* Supremo

A palavra Supremo em geral assume o caráter de adjetivo. No entanto, no direito brasileiro é usada em muitas ocasiões como sinônimo do Supremo Tribunal Federal (“STF”). Assim, nesse contexto assumiria o caráter de substantivo.

A Wordnet já mapeou o adjetivo “supremo”.

|  |
| --- |
| [00792641-a](http://wnpt.brlcloud.com/wn/synset?id=00792641-a) supreme, sovereign **| soberano, supremo** |

Porém, a sua interpretação da instituição STF ainda não existe. Dessa forma, tendo em vista a diferença de classificação gramatical, podem ocorrer erros nas análises que usam a Wordnet como base.

1. Erro do Freeling

* Correta/Correto

Ambas as palavras, quando passadas no freeling são mapeadas para “corretar”. No entanto, (a) “corretar” não existe na Wordnet e (b) a palavra “correto” já havia sido corretamente mapeada.

|  |
| --- |
| [00631391-a](http://wnpt.brlcloud.com/wn/synset?id=00631391-a) right, correct **| certo, correto, justo** |

* Estados

“Estados”, no freeling, está mapeado para “estados”. Esse mapeamento possui dois problemas. Em primeiro lugar, o freeling não encontrou um *synset* para “estados”. Em segundo, ele mapeou como Nome Próprio, pois a palavra “Estados” estava com letra maiúscula.

Em geral, “estados” é um nome comum. No entanto, nesse caso é um sinônimo para a entidade do Governo de um País (ex: o “Estado brasileiro”/ o “Estado não é legítimo”).

1. Triagem Inexistente

* Seccional
* Indenização

Esta palavra não possui mapeamento algum na Wordnet. Seu significado seria *“oferecer ou receber uma compensação ou reparação por um prejuízo ou dano sofrido”[[6]](#footnote-6).*

“Indenização” remete à reparação por determinado dano provocado por uma parte. No direito, a indenização tem o objetivo de fazer retornar a parte lesada ao estado inicial.

* Ministério Público

“Ministério Público” no Brasil é a denominação de uma instituição essencial à justiça. As palavras “Ministério” e “Público”, separadas, não representam a referida instituição, e os seus significados, mesmo unidos, também não seriam suficientes para a definir. Assim, seria necessário um *synset* próprio para “Ministério Público”.

Um bom mapeamento pode ser extraído da definição presente no art. 127 da Constituição Federal (“CF”):

“***Art. 127.*** *o Ministério Público é instituição permanente, essencial à função jurisdicional do Estado, incumbindo-lhe a defesa da ordem jurídica, do regime democrático e dos interesses sociais e individuais indisponíveis*”.

* Julgado

Julgado muitas vezes é dito como sinônimo de julgamento. Assim, apesar de não haver mapeamento na Wordnet para julgado, este seu significado poderia ser mapeado em conjunto com o seguinte *synset:*

|  |
| --- |
| [01191158-n](http://wnpt.brlcloud.com/wn/synset?id=01191158-n) ruling, opinion **| sentença, julgamento, ditame** |

Além disso, outra definição para “Julgado” é o particípio do verbo julgar. Essa é a definição achada na seguinte questão da OAB:

|  |
| --- |
| “ENUM\_Questão 77 Julgado dissídio coletivo entre uma categoria profissional e a patronal , em que foram concedidas algumas vantagens econômicas a a categoria de os empregados , estas não foram cumpridas de imediato por a empresa Alfa\_Ltda.” |

Como se pode depreender, nesse caso “julgado” quer dizer que o dissídio coletivo teve seu julgamento realizado, sendo mero particípio do verbo julgar, definido no seguinte *synset*:

|  |
| --- |
| [00672277-v](http://wnpt.brlcloud.com/wn/synset?id=00672277-v) judge **| julgar** |

* Cobrança

Não há mapeamento algum para esta palavra. A sua melhor definição é o substantivo (noun) do verbo (verb) “cobrar”. Assim, seria o “ato de cobrar”. Uma vez que o verbo “cobrar” já possui mapeamentos, grande parte das suas definições podem ser lincadas.

* Privativa
* Descumprimento

Embora a palavra “descumprimento” não esteja mapeada, o seu oposto, ou seja, a palavra “cumprimento” está. Assim, uma boa definição seria o oposto do seguinte *synset*:

|  |
| --- |
| [01127019-n](http://wnpt.brlcloud.com/wn/synset?id=01127019-n) enforcement **| coerção, cumprimento, cumprimento (legal), observação, aplicação, execução, imposição** |

Por conseguinte, ao invés de ser o ato de observar e obedecer a determinada coisa, é o contrário, qual seja, o ato de não observar e de desobedecer esta coisa.

* Concessionária

“Concessionária” é a denominação que se dá ao particular que recebeu a anteriormente referida “Concessão” da administração pública para prestar serviço público. Assim, o seu mapeamento deveria estar ligado ao da palavra “Concessão”.

No entanto, essa definição pode ser generalizada, não se limitando somente à esfera do direito. No fim, “concessionária” é aquele “*que recebeu a concessão, consentimento, para realizar algum serviço*”[[7]](#footnote-7).

* Constitucionalidade

No direito brasileiro, “constitucionalidade” significa que determinado ato está em conformidade com o texto e as normas da CF brasileira.

Na prova da OAB, a palavra apareceu no seguinte contexto:

|  |
| --- |
| “ENUM\_Questão 20 Em relação ao controle de constitucionalidade em face de a Constituição\_Estadual , assinale a alternativa correta” |

O controle de constitucionalidade mencionado é basicamente um controle realizado pelo Poder Judiciário para se averiguar se determinada norma ou ato está de acordo com a CF.

* Improbidade
* Habilitação
* Juntada
* Lide
* Arcar
* Vinculante
* Reintegração
* Sumaríssimo
* Sucumbência
* Ressarcimento
* Provimento
* Incabível
* Liminar
* Ltda

A palavra ltda, no direito brasileiro, é uma abreviação para “sociedade limitada”, uma das mais comuns formas de organização societária. Na Wordnet, essa palavra não foi mapeada. No entanto, “sociedade limitada” já possui uma interpretação corretamente feita.

|  |
| --- |
| [08185211-n](http://wnpt.brlcloud.com/wn/synset?id=08185211-n) Ld., limited\_company, Ltd. **| sociedade limitada** |

Note que a abreviação da palavra, em inglês, já está no banco de dados (Ltd.), faltando somente o equivalente em português.

**III - ANÁLISE QUALITATIVA**

As informações levantadas na seção anterior nos permitem apurar a eficácia do mapeamento da Wordnet no tocante às provas da OAB. Das palavras analisadas, algumas conclusões qualitativas podem ser extraídas:

1. Entre as [*número*] palavras com maior ocorrência, [--] tinham Interpretações Inexatas e [--] não tinham sido triadas. Isso nos dá uma precisão de [*percentual*].
2. A palavra relevante com maior ocorrência foi [--]. A partir de nossa análise, concluímos que ela se enquadra como [---].
3. A primeira palavra de maior ocorrência que apresenta algum problema é [--].
4. [---]

**IV - CONCLUSÃO**

[Se necessário]

Precisão é alta?

Tem muito a ser melhorado?

**APÊNDICE I**

**Relatório de Problemas Técnicos**

A realização do relatório não foi um processo integralmente tranquilo e agradável. Ao longo das nossas etapas preparatórias e metodológicas, alguns problemas técnicos dificultaram o trabalho analítico, conforme será relatado em sequência:

1. AntConc + FreeLing

Para viabilizar a contagem das palavras, foram feitos o download e o uso da ferramenta AntConc. No entanto, o seu manuseio não é tão simples e ágil para quem nunca o utilizou anteriormente.

Em primeiro lugar, a ferramenta demora consideravelmente para fazer um cálculo das palavras nos documentos. Tal demora é natural. Contudo, toda vez que se muda a aba interna do AntConc, o aplicativo congela e recomeça a contagem em algumas ocasiões.

Em segundo, a sua efetividade fica comprometida quando o AntConc é cruzado com os dados provenientes do FreeLing – ferramenta usada para realizar o mapeamento entre as palavras e a Wordnet. A forma como a ferramenta foi estruturada é extremamente não didática, e até o momento não conseguimos ranquear as palavras de maior ocorrência *junto com suas interpretações correspondentes na Wordnet.*

Toda vez que realizamos o ranqueamento, o output da Wordnet não é ranqueado ao lado das palavras. Assim, embora tenhamos as maiores ocorrências, não possuímos suas interpretações na base de dados digital. Consequentemente, as nossas análises ficam comprometidas, tendo em vista que precisamos abrir e procurar manualmente pelos termos na Wordnet, o que não somente torna o trabalho mais lento, mas também extingue a potencial análise da eficiência do algoritmo que correlaciona cada palavra com uma possível interpretação da base.

1. [--]

**APÊNDICE II**

**Códigos**

1. Código AWK para selecionar as palavras com mais recorrência que não estavam mapeadas na wordnet
2. Código em python que gera as frases nas quais as palavras são encontradas nas provas da OAB

1. https://www.significados.com.br/juizo/ [↑](#footnote-ref-1)
2. http://www.direitonet.com.br/dicionario/exibir/936/Recurso [↑](#footnote-ref-2)
3. http://www.elfez.com.br/elfez/execucao.html [↑](#footnote-ref-3)
4. http://www.direitonet.com.br/dicionario/exibir/827/Concessao-de-servico-publico [↑](#footnote-ref-4)
5. http://www.direitonet.com.br/dicionario/exibir/959/Nulidade-Novo-CPC-Lei-no-13105-15 [↑](#footnote-ref-5)
6. https://www.dicio.com.br/indenizacao/ [↑](#footnote-ref-6)
7. https://www.dicio.com.br/concessionaria/ [↑](#footnote-ref-7)