**Projectbeschrijving Boontje**

**Doel van het project**

In het boek *over De Kapellekensbaan & Zomer te Ter-Muren van Louis Paul Boon* (1977) wijdt G.J. van Bork een hoofdstuk aan het taalgebruik van Boon in de eerstgenoemde roman. Van Bork stelt zelf dat het moeilijk is om vast te stellen wat taaleigen is aan Boon, maar hij doet wel een aantal uitspraken over de specifieke kenmerken van de taal in *De Kapellekensbaan.* De beschreven specifieke kenmerken werden door van Bork echter niet gekwantificeerd, maar eerder intuïtief opgemerkt. Het doel van dit project is om de beweringen van van Bork te controleren via computationeel onderzoek. Hiervoor zal een programma geschreven worden in Python. Om beweringen als ‘veel’ of ‘meer dan gemiddeld’ te kunnen beoordelen wordt het digitaal beschikbare werk van Gerard Walschap als referentiemateriaal in deze studie betrokken. Eens de beweringen van van Bork getest zijn, weten we welke kenmerken werkelijk specifiek gelden voor Boons stijl. Vervolgens kunnen deze kenmerken getest worden op een aantal auteurs die Boons stijl heten na te volgen. Deze auteurs, o.a. Hugo Claus, Hugo Raes, Tom Lanoye en Rudi van Vlaenderen, kunnen dan gerangschikt worden op de mate waarin Boons specifieke stijl aanwezig is in hun eigen werk.

**Fases in het project**

1. *Corpusvoorbereiding*
   1. Corpus Boon
      1. Inlezen ruw corpus
         1. Oplossing zoeken op internet om epub om te zetten naar ander, bewerkbaar formaat met behoud van opmaak.
         2. Corpus inlezen, bewerkbaar als “write”.
      2. Uitfilteren commentaren (bv. inhoudstafel, nawoorden)
         1. Twee keer een CorpusKB maken met volgend bereik:
            1. Eerste zin KB:   
               het boek over de kapellekensbaan   
                 
               dat het boek is over de kinderjaren van ondineke, die geboren werd in tjaar 1800-en-zoveel... en die verliefd werd op meneer achilles derenancourt, directeur van de garenfabriek de filature, maar die op het laatste van het boek trouwen zal met het zielige oscarke...   
                 
               Wat ervoor komt zijn gegevens van de uitgeverij en mag weggegooid worden voor aanmaak CorpusKB.
            2. Laatste zin KB:  
               Maar ziet ge het? die stervende vrouw die een arm maar schoon meisken is geweest, met aan haar eigen sterfbed den doktoor en den pastoor die haar man en haar zoon moesten zijn geweest... dat zou een boek zijn, maar uw boek zal weeral geen boek zijn, er zal niets in staan over het leven gelijk het leven is.  
                 
               Wat erna komt is tekstverantwoording en mag weggegooid worden voor aanmaak CorpusKB.
      3. 3 bijkomende corpora maken uit een van de twee exemplaren CorpusKB om de 3 verhaallijnen binnen CorpusKB te onderscheiden:
         1. Elk stukje begint met:  
            1 of meer witregels + tab + titel in romein of cursief + 1 of meer witregels
         2. Voor CorpusVandaag volgt dan:  
            tekst in romein, mag onderbroken zijn met tabs
         3. Voor CorpusOndineke volgt dan:  
            tekst in cursief, mag onderbroeken zijn met tabs
         4. Voor CorpusReinaert volgt dan:  
            tekst in romein of cursief, mag onderbroken zijn met tabs + laatste stuk zin moet exact zijn:  
            . johan janssens
      4. Controleren of CorpusVandaag+CorpusOndineke+CorpusReinaert samen evenveel woorden bevatten als CorpusKB
      5. Inlezen bewerkte teksten om nu in totaal 4 corpora van Boon te hebben: CorpusKB, CorpusVandaag, CorpusOndineke, CorpusReinaert.
   2. Corpus Walschap
      1. Inlezen ruw corpus
         1. epub om te zetten naar ander, bewerkbaar formaat.
         2. Corpus inlezen, bewerkbaar als “write”.
      2. Uitfilteren commentaren (bv. inhoudstafel, nawoorden)
         1. Twee keer een CorpusWalschap maken met volgend bereik:  
            Eerste woorden: Uit het naamloze volk, dat toen nog niet zelf zijn naam kon schrijven, werkte zich een man op tot welstand. Dat is lang geleden, niemand onzer heeft er gedenken van.

Laatste woorden: dat fijne raadsel zweeft, waarvoor het sterven van Iphigénie hem aandachtig gemaakt hed en dat ons allen boeien blijft in dit aardse leven.

* + 1. 1 van de 2 CorpusWalschap versnijden in 3 romans om 3 corpora te maken:
       1. Eerste woorden CorpusCelibaat: Uit het naamloze volk, dat toen nog niet zelf zijn naam kon schrijven, werkte zich een man op tot welstand. Dat is lang geleden, niemand onzer heeft er gedenken van.  
          Laatste woorden CorpusCelibaat: en het Heerken heeft nog lichtekes met hoofd en vinger de maat geslagen, vijf of zes keren, meer niet, hij was te moe.
       2. Eerste woorden CorpusMens: Zijn strijd heeft Thijs op de ouderdom van acht jaren ingezet. Met een kat.  
          Laatste woorden CorpusMens: Zijn zoon is nu op de hoeve, de jongste van de drie, een klein stil manneke. We zullen zien wat het met hem wordt.
       3. Eerst woorden CorpusHoutekiet: Het leven van Houtekiet onze stambader begint met zijn ontmoeting van Lien, want van al wat daaraan voorafging weten wij niets.

Laatste woorden CorpusHoutekiet: dat fijne raadsel zweeft, waarvoor het sterven van Iphigénie hem aandachtig gemaakt hed en dat ons allen boeien blijft in dit aardse leven.

* + 1. Controleren of CorpusCelibaat + CorpusMens + CorpusHoutekiet samen evenveel woorden bevatten als CorpusWalschap.
    2. Inlezen bewerkte teksten om nu in totaal 4 corpora van Walschap te hebben: CorpusCelibaat, CorpusMens, CorpusHoutekiet, CorpusWalschap.
  1. Corpus Volgelingen (laatste fase, enkel indien nog ruimte)
     1. Verzamelen teksten
        1. CorpusClaus:  
           *Het verdriet van België*. In mijn bezit als epub.
        2. CorpusRaes:  
           *De vadsige koningen*. Te ontlenen in bib Boom en scannen met OCR.
        3. CorpusVlaenderen:  
           *Allemaal echt gebeurd.* Te scannen in erfgoedbibliotheek met OCR.
        4. CorpusLanoye:  
           *Gespleten en bescheten*. Te scannen in erfgoedbibliotheek met OCR.
     2. Inlezen ruw corpus
     3. Uitfilteren commentaren (bv. inhoudstafel, nawoorden)
     4. Inlezen bewerkt corpus

1. *Testen stijlkenmerken Boon*
2. *Presentatie resultaten*
3. *Testen stijlkenmerken Volgelingen* (laatste fase, enkel indien nog ruimte)
4. *Verwerking Volgelingen in presentatie resultaten* (laatste fase, enkel indien nog ruimte)