

## Keurig netjes 2: Een corpusstudie

Ton van der Wouden

Het Nederlands is een van de best beschreven talen van de wereld. En het is ook moeilijk om niet onder de indruk te raken van alle kennis en informatie die er over onze taal te vinden is in het [Taalportaal](#), de (digitale) [Algemene Nederlandse Spraakkunst](#) (ANS) en de [grote woordenboeken van het INT](#). Toch zijn er ook allerlei aspecten van het Nederlands waar we nog veel te weinig van afweten, maar die kennelijk deel uitmaken van de taalkennis van de moedertaalspreker. In de [vorige aflevering](#) demonstreerde ik dat aan de hand van de vaste verbinding *keurig netjes*: dat is een vaste verbinding die in Nederland veel vaker gebruikt wordt dan in België. Het zoeken naar dat soort combinaties in Van Dale en in de grote wetenschappelijke woordenboeken bleek evenwel niet altijd resultaat op te leveren, en in de grammatica's of taalcursussen vind je die informatie ook al niet. Arme tweede-taalleerders van het Nederlands.

Al doende bedacht ik me toen dat *keurig* met weinig andere bijvoeglijke naamwoorden dan *netjes* te combineren is, en aan die kwestie wil ik in deze aflevering aandacht besteden. Andere woorden en uitdrukkingen die wat mij betreft versterkt kunnen worden door *keurig* zijn de voltooid deelwoorden *opgeruimd*, *aangeharkt*, *opgevouwen* en *gekamd*, en voorzetselpredicaten als *in orde*, *op tijd* en *op z'n plek*, maar daarmee houdt mijn fantasie wel zo'n beetje op. \**Keurig doof*, \**keurig geslagen* en \**keurig met mate* klinken allemaal buitengewoon on-idiomatisch en kun je niet gebruiken om een hoge of keurige mate van respectievelijk *doof*, *geslagen*, en *met mate* aan te geven.

Tot zover mijn intuïtie. Om na te gaan of die klopt, ben ik in corpora gaan kijken. Ik ben begonnen met het [Corpus Gesproken Nederlands](#) (CGN), dat in het jaar 2004 voltooid is en zo'n 9 miljoen woorden bevat, waarvan ongeveer twee derde uit Nederland en een derde uit Vlaanderen. In dat corpus komt het woord *keurig* 161 keer voor, 145 keer in het Nederlandse deel en 16 keer in het Belgische (dat bevestigt de bevindingen van de vorige keer). In 11 van de 161 gevallen in het corpus wordt *keurig* gevolgd door *netjes* en in 1 geval door *nette*; al deze combinaties zijn afkomstig van Nederlandse sprekers. Volgens het (gratis) concordantieprogramma [antconc](#) is het woord dat (in het hele corpus) het vaakst onmiddellijk na *keurig* staat het voorzetsel *in* (12 keer, waarvan 3 keer *keurig in 't midden*), gevolgd door *netjes* (11 keer) en *op* (10 keer, waarvan 4 keer *op tijd*). De woorden die (in het hele corpus) het vaakst onmiddellijk vóór *netjes* staan zijn *heel* (55 keer), *niet* (26 keer), *wel* (25 keer) en *zo* (15). *Keurig* staat met 11 keer op de zevende plaats. Deze cijfers suggereren dat de kans dat een willekeurig voorkomen van *keurig* gevolgd wordt door *netjes* groter is dan andersom, met andere woorden, *keurig* is een betere voorspeller voor *netjes* dan *netjes* is voor *keurig*.

Hierboven noteerde ik dat de voltooid deelwoorden *opgeruimd*, *aangeharkt*, *opgevouwen* en *gekamd* misschien ook wel regelmatig na *keurig* zouden kunnen voorkomen. Dat wordt enigszins bevestigd voor *opgevouwen*, dat 2 keer na *keurig* voorkomt, en door *gekamd*, waarvoor dat 1 keer geldt, maar de combinaties *keurig opgeruimd* en *aangeharkt* komen in het hele CGN niet voor. Dat kan komen doordat mijn intuïties niet kloppen, of omdat het gebruikte corpus te klein is. Laten we, om daarachter te komen, overstappen naar een groter corpus. Ik neem daarvoor het [twittercorpus van Gosse Bouma](#), omdat het lekker groot is (gebaseerd op miljarden tweets) en een mooie zoekinterface heeft. In dat corpus komt *keurig* maar liefst 115757 keer voor. De woorden (en leestekens) die het vaakst onmiddellijk op *keurig* volgen staan in het onderstaande tabelletje:

10572	keurig op
5321	keurig in
4387	keurig uit
3597	keurig onderhouden

3507	keurig !
2844	keurig aan
2784	keurig .
2623	keurig ,
2588	keurig netjes
2023	keurig bij
1780	keurig gedragen

Ook hier is het niet *keurig netjes* dat bovenaan staat, maar *keurig op*, dat bij nadere beschouwing in ongeveer de helft van gevolgd blijkt te worden door *tijd*. Bij het daaropvolgende *keurig in* zien we een veel minder sterk effect: de meest frequente combinatie is *keurig in de* (waar dan weer met enige regelmaat nomina als *rij*, *kast*, *bus* en *hoek* op volgen), maar ook *keurig in orde*, *keurig in pak* en *keurig in bed* zijn interessant. De combinatie *keurig uit* daarentegen wordt bijna altijd gevolgd door een leesteken, dus dat zal vaak gaan om zinnen als *dat ziet er keurig uit*. De meest voorkomende adjectieven na *keurig* zijn volgens deze telling *onderhouden*, *netjes* en *gedragen*; verderop in de statistieken vinden we nog *afgewerkte* (720) en *nette* (633).

Een vergelijkbaar, maar niet geheel identiek beeld vinden we in het (nog grotere) [Google-corpus](#), dat eveneens toegankelijk gemaakt is door Gosse Bouma, en een soortgelijke zoekinterface biedt:

43062	keurig onderhouden
36079	keurig op
25852	keurig in
18597	keurig uit
18580	keurig netjes
14475	keurig en
13002	keurig verzorgd
9870	keurig aan
9349	keurig ingericht
9220	keurig verzorgde
7541	keurig de
7457	keurig afgewerkt
7029	keurig nette
6756	keurig afgewerkte
6086	keurig een
5982	keurig met
5853	keurig aangelegde
5703	keurig hotel
5298	keurig bij
4718	keurig schoon
4585	keurig verpakt
4493	keurig voor
3401	keurig ingerichte

In deze lijst staat *keurig onderhouden* helemaal bovenaan, en staat *keurig netjes* wat hoger dan in de twitterlijst. De verschillen, die in mijn ogen niet dramatisch zijn, zijn wellicht te verklaren uit het feit dat zowel *keurig* als *netjes* misschien inmiddels toch een beetje een oubollig woord is, dat niet zo past in een "hip" medium als twitter. Het internet als geheel, zoals toegankelijk via Google, bevat waarschijnlijk tekstmateriaal van meer verschillende aard.

Ook hier zien we trouwens weer veel voorzetsels bovenaan de lijst, met *op* als lijstaanvoerder. De combinatie *keurig op* wordt vaak gevolgd door nomina als *tijd*, *koers* en *schema*, maar vooral door lidwoorden (in afnemende volgorde *op de* met als belangrijkste zelfstandig naamwoord *achtergrond*, *op een* vooral gevolgd door *rijtje*, en *op het* gevolgd door *gemiddelde*) en de bezittelijke voornaamwoorden *zijn* (*keurig op zijn plaats*, *beurt*, *plek*) en *hun*.

Als we trouwens zoeken naar woorden die vaak onmiddellijk vóór *netjes* staan, dan staan *weer*, *wel*, *zo niet*, en *staat* bovenaan in de twitterdata. De versterker *keurig* vinden we daarin pas op de 51<sup>e</sup> plaats: dat bevestigt onze eerdere constatering op basis van het CGN dat *keurig* een goede voorspeller is voor *netjes*, maar *netjes* niet zo'n goede voorspeller voor *keurig*. In het google-corpus staat *keurig* trouwens aanzienlijk hoger, namelijk op de 21<sup>e</sup> plek. Maar ook in deze lijst is *keurig* niet de meest voorkomende versterker van *netjes*: boven dat woord staan (in afnemende volgorde) *heel*, *zeer*, en *erg*, die een veel algemenere distributie kennen.

Laten we tot slot nog eens de blik richten op alles wat vaak onmiddellijk ná *keurig* worden aangetroffen: dat zijn (in alfabetische volgorde) adjectieven en predicaten als *aan de regels*, *aangelegd(e)*, *afgewerkt(e)*, *gedragen*, *in orde*, *ingericht(e)*, *netjes*, *onderhouden*, *op de achtergrond*, *op een rijtje*, *op koers*, *op schema*, *op tijd*, *op zijn plaats*, en *verzorgd(e)*. Dat zijn allemaal woorden en uitdrukkingen die verwijzen naar eigenschappen die in onze maatschappij positief worden gewaardeerd.

Dat *keurig* de functie van adjectiefversterker alleen lijkt te kunnen vervullen met een handjevol positieve adjectieven en predicaten is relevante kennis, waar bijvoorbeeld tweede-taalleerders zeer gebaat bij zouden zijn. Moedertaalsprekers hebben daar geen kennelijk geen moeite mee, die halen het niet in hun hoofd om combinaties te vormen als *\*keurig doof*, *\*keurig geslagen* en *\*keurig met mate*. Kennelijk is dit deel van hun kennis van het Nederlands. Maar je vindt die informatie niet in de woordenboeken, niet in de grammatica's, en niet in de leerboeken Nederlands als tweede taal.

### Literatuur

Gosse Bouma. N-grams Frequencies for Dutch Twitter Data ([pdf](#)), *CLIN Journal*, volume 5 (2015).