

De digitale Pallas, een 18e-eeuws meertalig Russisch woordenboek met 60.000 taaldata

Verslag van lopend onderzoek

Nicoline van der Sijs, INT Leiden en Radboud Universiteit Nijmegen

Een team van onderzoekers werkt momenteel aan een geannoteerde digitale editie van het achttiende-eeuwse Russische vergelijkende woordenboek van alle talen en dialecten, samengesteld door de Pruisische geleerde Peter Simon Pallas op initiatief van niemand minder dan de Russische keizerin Catharina de Grote. Dit woordenboek bevat taaldata van honderden verschillende talen, waarvan sommige inmiddels uitgestorven zijn en andere gevaar lopen uit te sterven. Voor sommige talen zijn de gegevens in het woordenboek de oudste bron of een van de oudste bekende bronnen. Hoe is dit woordenboek tot stand gekomen?

In 1784 stortte keizerin Catharina zich op de studie van talen. Zij maakte een lijst van zo'n driehonderd Russische begrippen en liet deze vertalen in alle talen en dialecten die zij kon vinden. Ze begon zelfs eigenhandig een woordenboek samen te stellen op basis van het verzamelde materiaal, te beginnen met 'Caraïbische' woorden; zie figuur 1. In haar ogen was dit woordenboek 'misschien wel het nuttigste onderzoek dat ooit is verricht op het gebied van alle talen en woordenboeken, en vooral van belang voor de Russische taal'.

In 1785 riep Catharina de hulp in van de Pruisische arts, zoöloog, botanicus, geograaf en ontdekkingsreiziger Peter Simon Pallas, die in 1767 in Russische dienst was getreden en lid was van de Russische Academie van Wetenschappen. Pallas was een internationaal vermaard geleerde met een grote talenkennis. Hij breidde de woordenlijst van Catharina uit en stelde een *Modèle du vocabulaire qui peut servir à la comparaison de toutes les langues* samen, met 443 begrippen in het Russisch, Duits, Latijn en Frans. Dit 'model' werd niet alleen toegezonden aan de gouverneurs van de provincies van het uitgestrekte Russische rijk, maar ook aan Russische diplomaten over de hele wereld, en het werd uitgedeeld aan buitenlandse diplomaten in Rusland. Alle ontvangers kregen het verzoek te zorgen voor vertalingen van de begrippen in zoveel mogelijk talen. Dit leidde ertoe dat Pallas een groot aantal taaldata ontving, dat hij zelf nog vermeerderde met behulp van gegevens uit gedrukte woordenboeken.

Al in 1787 werd het eerste deel van de *Linguarum Totius Orbis Vocabularia Comparativa Augustissimae Cura Collecta*, of in het Russisch Сравнительные словари всѣхъ языковъ и нарѣчій собраные десницею всевысочайшей особы gepubliceerd, anoniem maar met een voorwoord van Pallas. Het tweede deel van deze 'Vergelijkende woordenlijsten van alle talen en dialecten in de wereld, verzameld dankzij de zorg van de allerhoogste persoon' kwam uit in 1789. De twee delen bevatten 273 begrippen met hun vertaling in 200 genummerde talen, eindigend met de namen van de telwoorden 1-10, 100 en 1000 in 222 talen. Het eerste begrip was Богъ (God), dat als voorbeeld kan dienen voor de eenvoudige opzet van het woordenboek (tussen vierkante haken heb ik de vertaling toegevoegd):

Богъ (God)

1 По Славянски – Богъ. [1 Slavisch Bog]

2 По Славяно-Венгерски – Бугъ. [2 Slavisch-Hongaars Bug]

3 По Иллирійски – Боогъ. [3 Illyrisch Boog]

...

199 На островахъ Маркезанскихъ – Етуо. [199 Op the Marquesaseilanden Etuo]

200 На островахъ Сандвича – Итуа. [200 Op the Sandwicheilanden Itua]

Na de publicatie van het woordenboek bleven nieuwe taaldata binnenkomen. Daarom besloot Catharina dat er een nieuwe editie moest komen, met een nieuwe structuur: zij gaf opdracht de woorden in de verschillende talen alfabetisch te ordenen. De nieuwe editie werd gepubliceerd in 1790-1791 in vier delen en met een licht gewijzigde titel: Сравнительный словарь всѣхъ языковъ и нарѣчій, по азбучному порядку расположенный, of Vergelijkend woordenboek van alle talen en dialecten, in alfabetische volgorde. Voorwoord en naam van de redacteur ontbreken, maar Friedrich Adelung onthult in zijn omvangrijke publicatie over het woordenboek, getiteld *Catherinens der Grossen Verdienste um die vergleichende Sprachenkunde* (1815), dat Theodor Jankiewitsch de Miriewo, directeur van de Pedagogische Academie, de redactie op zich had genomen omdat Pallas andere bezigheden had. Het woordenboek is opgezet in drie kolommen (woord, begrip en taal), en begint als volgt:

| | | |
|---|-----|--|
| А | Кто | По Ирландски [а, wie, Iers] |
| А | Онъ | По Сарамакски въ Суринамѣ [а, hij, Saramakkaans in Suriname] |
| А | Онъ | Суринамски Креольски [а, hij, Surinaams creools] |
| А | Она | Суринамски Креольски [а, zij, Surinaams creools] |
| А | Да | Татарски около Кузнецка [а, ja, Tataars rond Koeznetsk] |

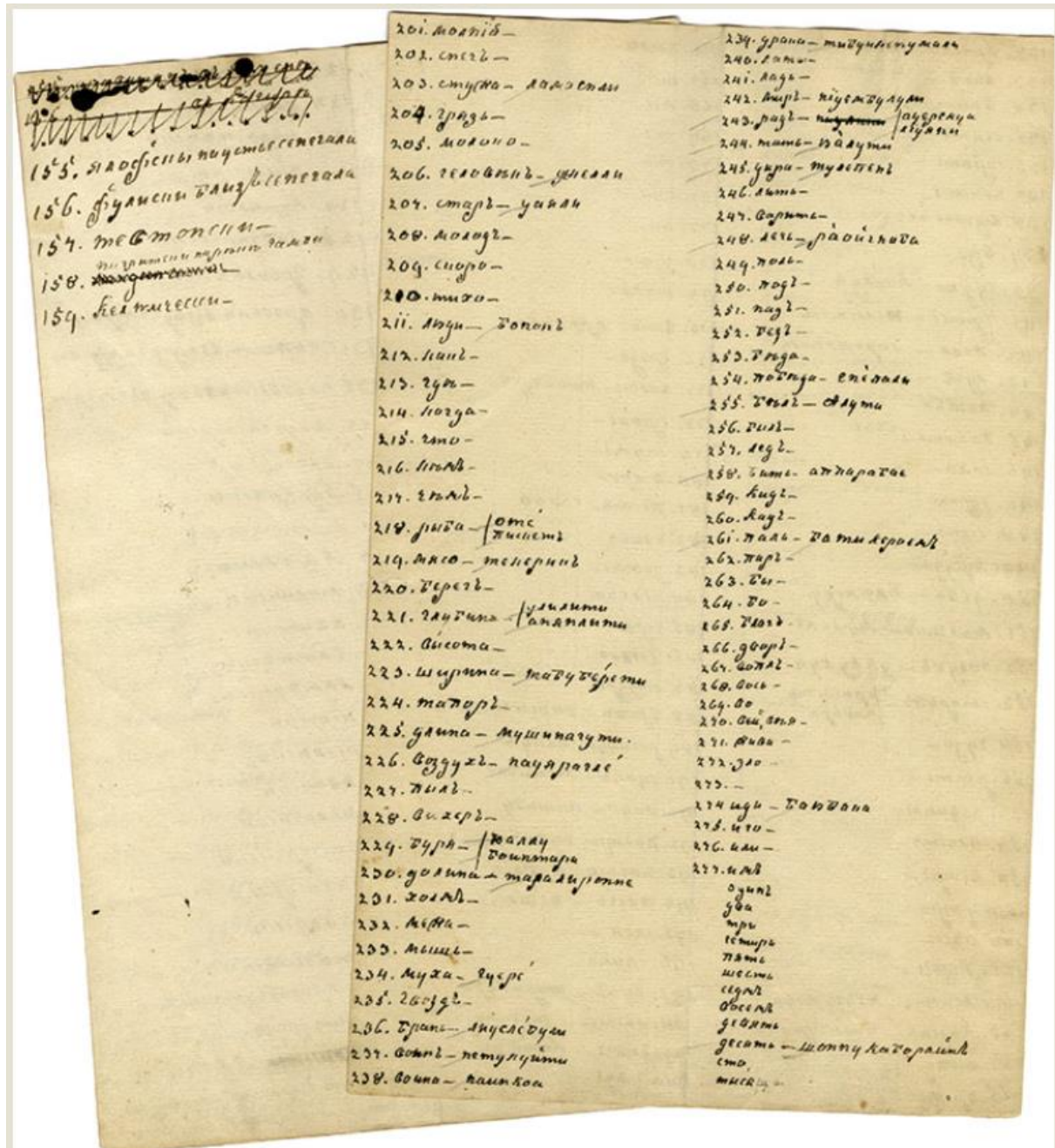
Slechts weinig geleerden buiten Rusland zijn bekend met het woordenboek, omdat het in het Russisch en het Cyrillisch is geschreven. Maar ook binnen Rusland is het grotendeels onbekend, omdat de meeste van de duizend exemplaren van de tweede druk onmiddellijk verdwenen in het keizerlijk archief. Deze onbekendheid is betreurenswaardig, want het woordenboek bevat een schat aan taaldata. Daarom heeft een team van onderzoekers in Nederland in 2020 het digitaliseren en annoteren van de tweede editie van het Pallas-woordenboek op zich genomen. Het Instituut voor de Nederlandse Taal in Leiden heeft voor ‘De digitale Pallas’ de Lexicon Interactive Tool Lex’it ter beschikking gesteld. Met deze tool zijn alle gegevens van het woordenboek inmiddels ingevoerd in een database.

Dankzij deze database weten we nu dat het woordenboek 61.960 woorden bevat voor 296 begrippen in 328 verschillende talen – dat is veel meer dan tot nu toe bekend was. Tot die talen behoort ook een groot aantal inheemse talen in Noord- en Zuid-Amerika en Afrika. We vonden opmerkelijke verschillen tussen het aantal woorden per taal: voor sommige talen vermeldt het woordenboek een groot aantal synoniemen, bijvoorbeeld voor het Japans (859 woorden voor de 296 begrippen). Voor andere talen is het aantal woorden echter zeer gering (23 woorden voor het Singalees, slechts 2 woorden voor het Thai).

Onze volgende taak is het toevoegen van transcripties en de oorspronkelijke (niet-Cyrillische) spelling aan de woorden in de verschillende talen, en het toevoegen van de moderne Engelse namen aan de taalnamen en taalclassificatie van Pallas. Tijdens dit annotatieproces stuiten we op allerlei onregelmatigheden. Zo zijn woorden soms ondergebracht bij een verkeerde taal (*Wanst* ‘buik’ is Duits, niet Nederlands zoals Pallas beweert), en zijn begrippen soms foutief vertaald (Russisch шум ‘lawaaï’ is niet hetzelfde als Nederlands *alarm*).

Als de annotatie is afgerond, zal de geannoteerde database op een openbare website worden gepubliceerd, zodat het materiaal gebruikt kan worden voor historisch taalkundig, taaltypologisch, lexicostatistisch en colexificatie-onderzoek. Verder zijn we van plan een boek samen te stellen waarin verschillende auteurs bijdragen leveren over de historische achtergrond van het Pallas-woordenboek en over de verschillende talen en taalfamilies die in het woordenboek zijn opgenomen. Iedereen die geïnteresseerd is in het project of meer informatie wil, kan contact opnemen met Nicoline van der Sijs: post@nicolinevdsijs.nl.

Het project 'De digitale Pallas' wordt gecoördineerd door Tjeerd de Graaf, Wim Honselaar, Janine Jager, Bruno Naarden en Nicoline van der Sijs. Medewerkers zijn Melle Groen, Marien Jacobs, Richard Kellermann Deibel, Martijn Knapen, Djoeke Leguijt, Sasha Lubotsky, Michael Nestorowysch, Tamara Schermer en Vincent Wintermans.



Figuur 1. Een pagina van de woordenlijst van Catharina de Grote waaraan ze de vertaling in het 'Caraiëisch' heeft toegevoegd.