

Factoren bij het herkennen van cognaten in onbekende talen: algemeen of taalspecifiek?

Jan Vanhove en Raphael Berthele

T&T 65 (2): 171–210
DOI: 10.5117/TET2013.2.VANH

Abstract¹

The correct recognition of cognate relationships is a precondition for receptive multilingualism within language families. Several item- and participant-related factors have been identified that affect the probability with which readers and listeners correctly recognise and translate cognates in an unknown but genealogically related foreign language. Most of the relevant findings, however, are based on tasks featuring cognates in only one unknown language, leaving open the possibility that these factors play a (somewhat) different role depending on the specific stimulus language. To address this issue, 98 German-speaking participants were asked to translate 181 written cognates from four Germanic languages (Danish, Dutch, Frisian and Swedish). The effects of stimulus similarity to known (L1, L2, ..., Ln) cognates, cognate frequency and the participants' foreign language skills on translation accuracy were assessed and found not to vary substantially between stimulus languages. This indicates that the validity of studies on cognate recognition extends beyond the language from which the stimuli were sampled. Secondary findings are discussed as well.

Keywords: cognate recognition, cross-linguistic similarity, Germanic languages, intelligibility, Levenshtein distance, receptive multilingualism

1. Inleiding

Onder cognaten worden woorden uit verschillende taalvariëteiten verstaan die een gemeenschappelijke etymologische oorsprong hebben, waardoor ze vaak een zekere formele gelijkenis vertonen. Cognaten zijn onontbeerlijk

voor de glottochronologie (bv. Dyen, Kruskal en Black, 1992; Gray en Atkinson, 2003) en de manier waarop meertaligen cognaten verwerken levert belangrijke inzichten op in hoe het mentale lexicon georganiseerd is (bv. Costa, Caramazza en Sebastián-Gallés, 2000; Lemhöfer, Dijkstra, Schriefers, Baayen, Grainger en Zwitserlood, 2008; Van Assche, Duyck, Hartsuiker en Diependaele, 2009; Van Hell en Dijkstra, 2002). Vanuit toegepast oogpunt zijn cognaten interessant omdat ze voor bepaalde L1-L2-combinaties (zoals Spaans-Engels) toegang kunnen verschaffen tot een groot receptief reservoir aan infrequente, gespecialiseerde L2-woorden (Bravo, Hiebert en Pearson, 2007; Lubliner en Hiebert, 2011) en ook productief makkelijker te leren zijn (Lotto en De Groot, 1998; De Groot en Keijzer, 2000). De correcte identificatie van cognaatverbanden vormt bovendien een hoeksteen voor receptieve meertaligheid op basis van taalverwantschappen (Möller en Zeevaert, 2010; Van Heuven, 2008). Dit is een soort talige constellatie waarin lezers of luisteraars (ten dele) kunnen begrijpen wat er in een bepaalde vreemde taal (Lx) geschreven of gezegd wordt zonder ooit deze taal geleerd te hebben en dit doordat deze taal een zekere gelijkenis vertoont met een taal die ze wel beheersen (L1, L2, ..., Ln).

Wanneer we het over talen hebben, bedoelen we daarmee niet alleen standaardtalen. Voor ons omhelst het begrip meertaligheid ook de individuele competentie in diasystemen binnen gestandaardiseerde talen, zoals bijvoorbeeld het beheersen van zowel een Limburgs dialect als Standaardnederlands of van een Alemannisch dialect en het Standaardduits. Dialectsprekers zijn in dit opzicht – net als sprekers met een gedifferentieerd talig repertoire (met inbegrip van registers en sociolectische vaardigheid) – productief minstens twee- en receptief meertalig (De Bot, 1992; Grosjean, 2010; Orioles, 2004; Wandruszka, 1981). Ze beschikken in het bijzonder over de vaardigheid systematische intertalige correspondenties in hun diasysteem te herkennen. Deze vaardigheid stelt hen dan weer in staat zich makkelijker toegang te verschaffen tot verwante vreemde talen, specifiek via cognaten (Berthele, 2008). Onderzoek naar cognatherkenning in verwante vreemde talen is dus intrinsiek verbonden met vraagstellingen uit de variatielinguïstiek: Cognate woorden in verschillende taalvariëteiten zijn dagelijkse kost voor dialectsprekers, die ze moeten leren herkennen, verstaan en vaak ook produceren. Zulk onderzoek verruimt louter de blik tot variëteiten van het ‘Germaans’ (i.p.v. het Nederlands of het Duits). Hierbij is het doel niet om de diasystemen zelf te beschrijven, maar gaat het om de spontane interpretatieprocessen die zich in de hoofden van meertaligen afspelen wanneer ze met enigszins vertrouwde woordvormen geconfronteerd worden.

Correcte cognaatherkenning in een Lx is geen vanzelfsprekendheid. Een respectabele hoeveelheid onderzoek is bijgevolg verricht naar factoren die Lx-cognaatherkenning en receptieve meertaligheid in het algemeen gunstig dan wel nefast kunnen beïnvloeden. Zulk onderzoek heeft enerzijds tot doel na te gaan waarom sommige cognaten makkelijker te herkennen zijn dan andere en anderzijds waarom bepaalde lezers of luisteraars vaardiger zijn in het herkennen van cognaten dan anderen. Voor beide doelstellingen wordt vaak gebruik gemaakt van taken waarin proefpersonen worden gevraagd gesproken of geschreven Lx-cognaten naar hun moedertaal te vertalen. Het percentage correcte vertaalpogingen per cognaat drukt hierbij veelal de moeilijkheid of transparantie van het cognaat uit; als maat voor de individuele cognaatherkenningsvaardigheid wordt in de regel het percentage correcte vertaalpogingen per proefpersoon genomen.

De transparantie van Lx-cognaten kan tot op zekere hoogte gemodelleerd worden met behulp van de fonetische of orthografische overlap tussen het te vertalen woord en zijn cognaat in een beheerste taal. Op deze factor gaan we later dieper in. Daarnaast zijn er aanwijzingen dat in het bijzonder discrepanties tussen de medeklinkers nefast zijn voor cognaatherkenning (Berthele, 2011; Gooskens, Heeringa en Beijering, 2008; Möller, 2011; Möller en Zeevaert, 2010), en dat lezers en luisteraars zich sterker laten leiden door het begin van het woord dan door het einde ervan (Berthele, 2011; Möller, 2011; Möller en Zeevaert, 2010).

De individuele vaardigheid Lx-cognaten correct te herkennen en te vertalen lijkt mee bepaald te zijn door reeds voorhanden talenkennis (hierover later meer) en de leeftijd van de proefpersonen: volwassenen verstaan geschreven cognaten in verwante talen beter dan kinderen en adolescenten, terwijl het begrip van gesproken cognaten piekt tussen 30 en 50 jaar en daarna weer afneemt (Vanhove en Berthele, aangenomen). Ook de mate van voorafgaand contact met de verwante taal speelt een rol: wie al ervaring met de Lx in kwestie heeft, verstaat cognaten in deze taal ook beter (Delsing en Lundin Åkesson, 2005; Jensen, 1989). Tot slot zouden attitudes tegenover de verwante taal kunnen meespelen: wie positief is ingesteld ten opzichte van de verwante taal, zou meer moeite doen om deze te verstaan, wat tot een beter cognaatbegrip leidt (Wolff, 1959). De rol van attitudes en hun eventuele causale impact op cognaatherkenning zijn echter omstreden (Gooskens en Van Bezooijen, 2006; Jensen, 1989; Schüppert en Gooskens, 2011; Van Bezooijen en Gooskens, 2005b).

Wanneer we echter op de basis van deze studies generalisaties willen maken met betrekking tot de herkenning van geïsoleerde cognaten in het algemeen (d.w.z. onafhankelijk van een specifieke Lx), dan stuiten we op

een potentieel probleem.² In de meeste studies kregen de proefpersonen namelijk Lx-cognaten uit slechts één bepaalde Lx te lezen of te horen. Dit maakt het moeilijk om te achterhalen of bijvoorbeeld voorhanden talenkennis en cognitieve factoren de cognaatherkenningsvaardigheid in het algemeen beïnvloeden of slechts in de specifieke Lx die onderzocht werd. Plausibeler is dat zulke factoren weliswaar een algemene rol spelen, maar dat ze van Lx tot Lx verschillend zwaar doorwegen. Zo kunnen we ons de vraag stellen of de rol van pakweg de fonologische of orthografische overlap tussen een Lx-stimulus en zijn L₁-, L₂-, ..., L_n-cognaten niet al naar gelang de specifieke Lx systematisch verschilt. Zulke systematische verschillen zouden zich kunnen voordoen als de proefpersonen minder geneigd zijn om hun L₁-, L₂-, ..., L_n-kennis aan te spreken bij de ene Lx dan bij de andere, bijvoorbeeld omdat de psychotypologische afstand (Kellerman, 1983) – d.w.z. het subjectieve aanvoelen van de verwantschap tussen talen door de proefpersoon, die niet noodzakelijk nauwkeurige kennis heeft van de daadwerkelijke genealogie – tot de ene Lx groter is dan tot de andere. Een alternatieve mogelijkheid is dat proefpersonen bepaalde (al dan niet terechte) voorstellingen hebben over de grafeem-foneem-correspondenties in de Lx'en: een Duitstalige die weet dat de Nederlandse *oe* als [u] en niet als [ø] wordt uitgesproken, maar niet weet dat de Zweedse *å* als [o] wordt uitgesproken, ondervindt mogelijk minder hinder van de orthografische discrepanties tussen *moed* en *Mut* dan van die tussen *månad* en *Monat*. Voor zover ons bekend is, werden in slechts drie studies (Berthele, 2008, 2011; Berthele en Lambelet, 2009) stimuli uit twee Lx'en gebruikt, maar het totale aantal stimuli was in deze studies vrij laag (28 tot 29 stimuli), waardoor ze geen robuuste generalisaties toelaten.

In deze studie onderzoeken we daarom de mate van generaliseerbaarheid van de resultaten van studies naar Lx-cognaatherkenning. Voor een goed begrip definiëren we cognaten als etymologisch verwante woorden die semantisch equivalent zijn (voor op zijn minst een betekenis bij polyseme woorden). Valse vrienden met een gemeenschappelijke etymologie, bv. het Engels-Duitse woordpaar *eventual-eventuell*, worden hier dus niet als cognaten beschouwd. We vroegen Duitstalige Zwitsers om in totaal 200 woorden te vertalen uit vier Germaanse talen die ze niet beheersten (Deens, Fries, Nederlands en Zweeds). Uit praktische overwegingen concentreerden we ons hierbij uitsluitend op cognaatherkenning in de geschreven modaliteit. Concreet stellen we in deze studie de vraag in welke mate vier uitgekozen predictoren interageren met de specifieke Lx waarin een stimulus gepresenteerd wordt bij het statistisch modelleren van cognaatherkenning. Zijn deze interacties bescheiden, dan kunnen we

aannemen dat deze predictoren een vergelijkbare rol spelen in verschillende Lx'en en dat de bevindingen van studies naar Lx-cognaatherkenning zeer vermoedelijk grotendeels veralgemeend kunnen worden. Zijn de interacties echter beduidend, dan kunnen we niet zonder meer aannemen dat zulke studies een groter geldingsbereik hebben dan de specifieke Lx-L1, L2, ..., Ln-combinatie die getest werd. Tegelijkertijd biedt deze studie ons de gelegenheid om waarnemingen uit vorig onderzoek aan een bijkomende onafhankelijke empirische verificatie te onderwerpen.

2. Geselecteerde variabelen

We selecteerden vier variabelen waarvan de invloed op correcte cognatherkenning al in eerdere studies onderzocht werd. Twee daarvan, nl. orthografische afstand en corpusfrequentie, zijn stimulusgerelateerd. De andere twee, t.w. aantal beheerste talen en kennis van het Engels, zijn persoonsgerelateerd. We stellen ze hier summier voor.

2.1 Orthografische afstand

Een voor de hand liggende voorspelling is dat Lx-stimuli die sterk lijken op hun L1-, L2-, ..., Ln-cognaten gemakkelijker en vaker juist geïdentificeerd kunnen worden. Een populaire methode om de graad van gelijkenis tussen cognaten kwantitatief te operationaliseren is het Levenshteinalgoritme (Levenshtein, 1966). Dit algoritme berekent het minimaal aantal invoegingen, deleties en substituties dat nodig is om een string in een andere om te vormen. Bij wijze van voorbeeld kan het Zweedse woord *riktning* ('richting') met een invoeging, een deletie en twee substituties worden omgevormd tot het Duitse *Richtung* (de hoofdletter buiten beschouwing gelaten, zie figuur 1). De (rauwe) Levenshteinafstand tussen *riktning* en *Richtung* bedraagt met andere woorden 4. Ook fonetische transcripties kunnen op deze manier worden vergeleken. Rauwe Levenshteinafstanden worden in de regel genormaliseerd zodat langere strings niet door de band genomen hogere Levenshteinwaarden hebben. Dit kan bijvoorbeeld door elke rauwe Levenshteinafstand te delen door de lengte van de langste string in de vergelijking.³ Bij *riktning-Richtung* zijn beide strings acht tekens lang en bedraagt de genormaliseerde Levenshteinafstand dus $4 / 8 = 0,5$. Levenshteinafstanden die op deze manier genormaliseerd worden, liggen altijd tussen 0 (totale orthografische of fonetische overlap) en 1 (geen orthografische of fonetische overlap).⁴ In wat volgt hebben we het steeds over genormaliseerde Levenshteinafstanden.

r	i	k		t	n	i	n	g	
r	i	c	h	t		u	n	g	
		S	I		D	S			= 4

Figuur 1 Voorbeeld van een Levenshteintransformatie. Het Zweedse riktning kan met behulp van een invoeging (I), twee substituties (S) en een deletie (D) tot het Duitse Richtung getransformeerd worden.

Het Levenshteinalgoritme wordt courant gebruikt in studies over receptieve meertaligheid en Lx-cognaatherkenning. Zo kan de verstaanbaarheid van gesproken teksten in een verwante taal in grote mate gemodelleerd worden aan de hand van hun fonetische Levenshteinafstand tot een L1-vertaling van de tekst (Beijering et al., 2008; Gooskens, 2007; Gooskens et al., 2008). Ook de verstaanbaarheid van geïsoleerde gesproken Lx-woorden kan ten dele voorspeld worden op basis van hun fonetische afstand tot een cognaat in de L1 (Doetjes en Gooskens, 2009; Gooskens, Kürschner en Van Bezooijen, 2011; Kürschner et al., 2008; Van Bezooijen en Gooskens, 2005a) of een andere beheerste taal (Berthele, 2011).

Er zijn indicaties dat ook de verstaanbaarheid van geschreven Lx-teksten tot op zekere hoogte samenhangt met hun Levenshteinafstand tot een L1-vertaling (Gooskens en Van Bezooijen, 2006; Van Bezooijen en Gooskens, 2005a,b). De rol van de Levenshteinafstand voor de verstaanbaarheid van geïsoleerde geschreven woorden – wat ons hier het meest interesseert – werd, voor zover ons bekend is, in slechts twee studies onderzocht. Berthele en Lambelet (2009) onderzochten de verstaanbaarheid van 29 Reto-Romaanse en Roemeense woorden voor Frans- en Italiaanstalige Zwitserse studenten. Het aantal correcte vertalingen per woord correleerde daarbij, weliswaar relatief zwak, met de orthografische Levenshteinafstand tussen de stimulus en zijn Frans of Italiaans cognaat ($r = -0,32$; de auteurs namen telkens de laagste van de twee Levenshteinafstanden). In een gelijkaardig onderzoek met 28 Deense en Zweedse werkwoorden bij Duitstalige Zwitsers vond Berthele (2011) geen correlatie tussen het aantal keer dat de stimuli correct vertaald werden en hun orthografische Levenshteinafstand tot hun cognaat in het Duits ($r = 0,14$), maar wel met hun Levenshteinafstand tot hun cognaat in het Engels ($r = -0,35$). Dit suggereert dat bij het operationaliseren van de orthografische overlap van een cognaat ook rekening gehouden moet

worden met cognaten in vreemde talen die de proefpersonen enigszins beheersen.

2.2 Corpusfrequentie

In de gesproken modaliteit kunnen hoogfrequente L1-woorden efficiënter verwerkt en geproduceerd worden (voor verwijzingen zie Luce en Pisoni, 1998: 2-4, en Norris en McQueen, 2008: 369). Zich baserend op zulke frequentie-effecten in de L1 stelt Van Heuven (2008) dat frequente woorden ook eerder voor de hand zullen liggen bij luisteraars die geconfronteerd worden met onduidelijke input. Concreet zou de waarschijnlijkheid op correcte Lx-cognaatherkenning bijgevolg groter zijn wanneer de stimulus een hoogfrequent L1-, L2-, ..., Ln-cognaat heeft dan wanneer het relevante cognaat eerder infrequent is. Dit idee werd empirisch getest door Kürschner et al. (2008), die de verstaanbaarheid van Zweedse woorden voor Denen linkten aan onder andere de frequentie van hun Deense cognaten. Cognaatfrequentie droeg weliswaar bij tot hun model, maar slechts minimaal.

Ook in de geschreven modaliteit kunnen hoogfrequente L1-woorden in de regel beter verwerkt worden dan infrequente woorden (zie Norris, 2006: 327, voor verwijzingen). Wij hebben echter geen weet van studies die cognaatfrequentie liëren aan de verstaanbaarheid van geschreven Lx-stimuli. Deze studie biedt dus een uitstekende gelegenheid om na te gaan in hoeverre er een verband bestaat en of dit verband van Lx tot Lx verschilt.

2.3 Aantal beheerste talen

Berthele (2011) en Berthele en Lambalet (2009) vonden indicaties dat het aantal talen dat iemand in meer of mindere mate beheerst positief (maar relatief zwak) correleert met de individuele Lx-cognaatherkenningsvaardigheid, dit zowel voor gesproken als voor geschreven stimuli. Een belangrijkere factor dan het aantal beheerste talen lijkt echter de aard van de beheerste talen te zijn: vooral lezers en luisteraars met goede kennis van twee of meer talen of taalvariëteiten die zowel met elkaar als met de Lx nauw verwant zijn, zijn in het voordeel (zie ook Berthele, 2008, en Gooskens et al., 2011, maar vergelijk Van Bezooijen, Gooskens en Kürschner, 2012). Dat meertaligen en in het bijzonder meertaligen in verwante talen in het voordeel zijn bij Lx-cognaatherkenning kan enerzijds verklaard worden doordat ze over meer potentiële transferroutes beschikken. Een bijkomende mogelijke verklaring, stelt Berthele (2011), is dat zulke meertaligen in staat zijn intertalige patronen (bv. grafeem- of foneemcorrespondenties) in hun lexicon te identificeren en deze speculatief toe te passen op andere

taalcombinaties. Deze vorm van geïnformeerd intertalgig speculeren noemt hij *abductie* (naar Peirce, 1931).

2.4 Kennis van het Engels

Zoals hierboven al kort vermeld draagt kennis van verwante talen bij tot een betere cognatherkenningsvaardigheid in een Lx. Onze proefpersonen zijn Duitstaligen en worden geconfronteerd met de hun onbekende Germaanse talen Deens, Fries, Nederlands en Zweeds. Bovendien mochten ze geen voorkennis hebben in een andere Germaanse taal, met één voor de hand liggende uitzondering: het Engels. Kennis van het Engels hangt bij Duitstaligen samen met de individuele Lx-cognatherkenningsvaardigheid (Berthele, 2011; Vanhove en Berthele, aangenomen), wat zowel verklaard kan worden door het hogere aantal transfermogelijkheden die extra kennis van het Engels met zich meebrengt (of de betere beschikbaarheid van zulke transfermogelijkheden) als door een verbeterd intertalgig inzicht dat bij abductieve processen (zie hierboven) van pas komt.

3. Methode

3.1 Proefpersonen

In de lente van 2012 werden in het kader van het werkcollege *Plurilinguale Kompetenz* aan de vakgroep voor Meertaligheid en Vreemdetalendidactiek aan de universiteit van Fribourg (Zwitserland) twaalf studenten ingeschakeld om proefpersonen te rekruteren, hen een vragenformulier te laten invullen en een vertaaltaak (zie hieronder) voor te leggen. De proefpersonen moesten minstens 16 jaar oud zijn en moesten een Zwitserduits dialect als hun moedertaal (of een van hun moedertalen) beschouwen. Bovendien mochten ze niet over voorkennis van Germaanse talen buiten het Standaardduits, Zwitserduitse dialecten en het Engels beschikken en mochten ze geen taalexperthen, zoals taalkundigen of tolken, zijn.

Alvorens aan de vertaaltaak te beginnen vulden de proefpersonen een kort vragenformulier in. Hierin moesten ze hun geslacht, leeftijd en moedertaal aangeven en moesten ze aanduiden of ze een taalgeoriënteerde opleiding hadden genoten. Bovendien werden ze gevraagd hun lees- en luistervaardigheid in het Engels, Frans en eventuele andere beheerste vreemde talen in te schatten aan de hand van het zelfbeoordelingsschema voor het Gemeenschappelijk Europees Referentiekader (CEFR)⁵. Het CEFR omvat

zes niveaus van A1 tot C2. Enkele proefpersonen gaven zelfbeoordelingen tussen twee niveaus in op (bv. 'B1-B2'). In zulke gevallen rondden we de beoordeling naar boven af. Deze niveaus worden in onze analyses numeriek gerepresenteerd (A1 = 1, A2 = 2, enz.).

In totaal werden 107 proefpersonen gerekruteerd. Drie onder hen gaven aan dat hun kennis van het Engels niet goed genoeg was om aan een van de zes CEFR-beschrijvingen te voldoen. Zes andere proefpersonen werden niet getest onder het directe toezicht van een student. Deze negen proefpersonen werden bijgevolg van onze analyse uitgesloten, waardoor er een steekproef van 98 proefpersonen overbleef. Wanneer de zes proefpersonen die niet onder direct toezicht getest werden toch in de analyses worden opgenomen, verandert er nauwelijks iets aan de bevindingen waarover hier gerapporteerd wordt; geïnteresseerde lezers kunnen dit zelf aan de hand van de online data en computercode uitrekenen.

De overblijvende 98 proefpersonen (33 mannen) varieerden in leeftijd van 16 tot 72 jaar (mediaan: 30 jaar). Ze waren als volgt opgedeeld naar moedertaaldialect: 26 uit de regio Zürich (incl. Aargau); 23 uit de regio Bazel; 20 uit Oost-Zwitserland (Appenzell, Sankt Gallen, Schaffhausen en Thurgau); 17 uit de regio Bern (incl. Solothurn); 14 uit Centraal-Zwitserland (Luzern, Schwyz en Zug).⁶ 70 proefpersonen gaven aan buiten Zwitserduitse dialecten en het Duits, Frans en Engels nog talen tot op zekere hoogte te beheersen. Het gaat hierbij om Italiaans (44), Spaans (38), Turks (5), Portugees (4), Russisch (4), Grieks (2), Albanees, Arabisch, Bulgaars, Kantonees, Koerdisch, Koreaans, Macedonisch, Pools, Reto-Romaans en Tsjechisch (allemaal 1).

3.2 Vertaaltaak

De vertaaltaak bestond uit vier bladzijden waarop telkens 50 substantieven (zonder context) in een Germaanse taal (Deens, Fries, Nederlands en Zweeds) gedrukt waren. 176 van deze woorden waren (in theorie) verstaanbaar via (onder meer) een Duits cognaat (bv. De. *opmærksomhed* 'aandacht' via Dt. *Aufmerksamkeit*), 86 via een Engels cognaat (bv. Zw. *sång* via Eng. *song*) en 58 via een Frans cognaat (bv. Zw. *elev* 'leerling' via Fra. *élève*). 90 woorden waren via meer dan een taal verstaanbaar. In totaal waren 181 woorden in theorie via een Duits, Engels of Frans cognaat verstaanbaar. Zij worden in wat volgt 'doelwoorden' genoemd. Deze doelwoorden werden geselecteerd uit een lijst met 384 frequente substantieven die door de onderzoeksgroep *Mutual intelligibility of closely related languages* (Rijksuniversiteit Groningen, PI: Charlotte Gooskens) werd samengesteld. Deze woordenlijst werd al in eerdere studies over de verstaanbaarheid

van gesproken cognaten gebruikt, onder andere door Kürschner et al. (2008), Kürschner en Gooskens (2011) en Gooskens et al. (2011). De precieze selectieprocedure wordt beschreven door Kürschner en Gooskens (2008): Uit het Corpus Gesproken Nederlands en uit de Nederlandse en Zweedse monologen in het Europarlocorpus werden telkens de 1.500 meest frequente woorden geselecteerd. In totaal ging het om 2.575 verschillende woorden, waarvan 815 substantieven. 24 Nederlandse scholieren van 15 tot 17 jaar oud werden gevraagd de substantieven aan te kruisen die ze niet kenden. 82 substantieven bleken niet door alle informanten gekend te zijn en werden uit de lijst verwijderd. 349 resterende substantieven werden willekeurig verwijderd, waardoor 384 substantieven overbleven. Deze werden naar de verschillende Germaanse talen vertaald.

Vier tot vijf substantieven per taal konden in principe niet zonder voorkennis van de verwante taal vertaald worden, bv. het Friese *heit* ('vader') en het Nederlandse *wet*. Zulke woorden worden hier 'profielwoorden' genoemd, een begrip uit de EuroCom-projecten (Stegmann en Klein, 1999).⁷ Proefpersonen die erin slaagden meer dan twee profielwoorden in een bepaalde taal te vertalen werden verondersteld te veel lexicale voorkennis in die taal te hebben en zouden van de analyses worden uitgesloten. Geen van de proefpersonen slaagde hier echter in. De volledige stimuluslijst met Duitse, Engelse en Franse cognaten is in de appendix te vinden.

De proefpersonen werden gevraagd te proberen de 200 woorden naar het Duits te vertalen. Daarbij werd hen verteld dat het uitsluitend om substantieven ging (met een korte lekenuitleg over wat substantieven zijn). Wanneer ze geen Duitse vertaling konden bedenken, mochten ze het woord onvertaald laten. Ongeveer de helft (50) van de proefpersonen kreeg eerst de Deense, vervolgens de Nederlandse en de Zweedse en tot slot de Friese woorden voorgeschoteld; de andere (48) proefpersonen werkten de taak in de omgekeerde volgorde af. Op die manier wilden we eventuele leer- of vermoeidheidseffecten neutraliseren. Op elke bladzijde stond duidelijk aangegeven uit welke taal de woorden op de bladzijde kwamen. Voor elke bladzijde hadden de proefpersonen maximaal een kwartier de tijd, wat ruimschoots voldoende bleek.

Elke vertaling werd door de studenten als juist of fout beoordeeld. Deze beoordelingen werden door de eerste auteur gecontroleerd en, zo nodig, verbeterd om een consistente beoordeling te garanderen.

3.3 Berekening van orthografische afstanden

We berekenen de orthografische afstanden tussen de stimuli en hun Duitse, Engelse en Franse cognaten met behulp van het Levenshteinalgoritme (zie 2.1). Voordat we de Levenshteinafstanden berekenden, werden alle strings naar kleine letters geconverteerd en van hun diakritische tekens ontdaan. Bovendien werden ligaturen, zoals het Deense *æ* en het Franse *œ*, in tweeën gesplitst.

Voor elk doelwoord werden er drie Levenshteinafstanden berekend, een voor elke brontaal (Duits, Engels en Frans). Wanneer er voor een doelwoord in een van de brontalen geen cognaat bestond, werd de betreffende Levenshteinafstand arbitrair als 1 (het maximum) gedefinieerd. Voor elk doelwoord berekenden we vervolgens ook de minimale Levenshteinafstand, d.w.z. de kleinste waarde van deze drie Levenshteinafstanden. Beschrijvende statistieken voor de taalspecifieke en minimale Levenshteinafstanden worden gegeven in tabel 1.

Tabel 1 Beschrijvende statistieken van de 181 doelwoorden.

Voor doelwoorden zonder cognaat in een bepaalde taal werd de Levenshteinafstand op 1 (het maximum) en de cognaatfrequentie op 0 (het minimum) gesteld. De cognaatfrequenties zijn log-getransformeerd ($\ln(\text{frequentie}+1)$).

		Spreidings- breedte	Mediaan	Gemiddelde	Standaard- afwijking
Levenshtein- afstand	Duits	0,00–1,00	0,40	0,41	0,20
	Engels	0,00–1,00	1,00	0,72	0,33
	Frans	0,00–1,00	1,00	0,81	0,31
	minimum	0,00–0,86	0,33	0,36	0,17
Cognaat- frequentie (log-getrans- formeerd)	Duits	0,00–7,05	3,08	3,11	1,57
	Engels	0,00–7,26	0,00	1,79	2,17
	Frans	0,00–6,27	0,00	1,18	1,89
	totaal	0,04–7,86	3,94	3,87	1,61

3.4 Extractie van cognaatfrequenties

Voor de Duitse en Engelse cognaten van de doelwoorden extraheerden we de corpusfrequenties per miljoen woorden uit SUBTLEX-DE (Brysbaert, Buchmeier, Conrad, Jacobs, Bölte en Böhl, 2011) en SUBTLEX-US (Brysbaert en New, 2009).⁸ De Franse cognaatfrequenties per miljoen woorden werden uit het vergelijkbare Lexique 3 (New, Brysbaert, Veronis en Pallier, 2007) geëxtraheerd.⁹ De lemma's in deze databases werden echter eerst geconverteerd naar kleine letters en ontdaan van hun diakritische tekens.

Zodoende werden ze op dezelfde manier voorbereid als de cognaten voor de Levenshteinberekeningen. De Duitse letter β komt in Zwitserse teksten nagenoeg niet voor en werd als *ss* herschreven. De frequenties van (eventueel nieuw verkregen) homografen werden daarna geaggregeerd. Een globale maat van cognaatfrequentie werd verkregen door de Duitse, Engelse en Franse cognaatfrequenties bij elkaar op te tellen. Zowel de taalspecifieke als de globale frequentiematen waren sterk rechts-scheef verdeeld, waardoor we een logaritmische transformatie uitvoerden. Omdat het logaritme van 0 niet gedefinieerd is, telden we eerst bij elke cognaatfrequentie 1 op, wat gebruikelijk is in studies over woordfrequenties: $\ln(\text{frequentie} + 1)$. Het zijn deze logaritmisch getransformeerde frequentiematen, waarvan de beschrijvende statistieken in tabel 1 staan, die we in onze analyses zullen gebruiken.

3.5 Statistische analyse

Om een antwoord te bieden op onze onderzoeksvraag maakten we gebruik van veralgemeende (logistische) lineaire modellen met gemengde effecten (*generalised linear mixed-effects models*; GLMM's) met de correctheid van elke individuele vertaalpoging als afhankelijke binaire variabele. We fitten deze modellen met behulp van het *lme4*-package (Bates, Maechler en Bolker, 2012) voor R (R Core Team, 2012). Om de resultaten van dit onderzoek tot in de details te begrijpen is vertrouwdheid met regressiemethoden noodzakelijk. Zowel bij het presenteren van de resultaten als bij de discussie bespreken we de modellen echter telkens ook in niet-technische bewoordingen. Hier lichten we informeel de bijzonderheden van GLMM's toe; voor iets formelere maar niettemin toegankelijke beschrijvingen verwijzen we naar Baayen (2008), Baayen, Davidson en Bates (2008) en Jaeger (2008).

In een logistisch GLMM wordt een binaire afhankelijke variabele (in ons geval *juist* versus *fout*) gemodelleerd in functie van zowel vaste effecten (*fixed effects*) als van toevalseffecten (*random effects*). Deze aanpak biedt het voordeel dat de afhankelijkheidsstructuur van de data gespecificeerd kan worden, waardoor de onafhankelijkheidsaannname van traditionele regressiemethoden omzeild kan worden. Interessant voor taalkundige en psycholinguïstische onderzoeken is verder dat trends in de data zowel op hun generaliseerbaarheid met betrekking tot nieuwe proefpersonen als met met betrekking tot nieuwe stimuli getest kunnen worden. Op deze manier kan de statistische valkuil van de *language-as-fixed-effect fallacy* (Clark, 1973; zie ook Coleman, 1964), de mogelijk foutieve veronderstelling dat een variabele voor alle talige stimuli hetzelfde effect heeft, vermeden worden. Hetzelfde geldt voor *Simpson's paradox* (Simpson, 1951; zie ook Jaeger, Graff,

Croft en Pontillo, 2011, en Schielzeth en Forstmeier, 2009), die de situatie beschrijft waarin het globale effect van een variabele tegengesteld is aan haar groepspecifieke effecten.

Vaste effecten zijn de effecten die men ook in gewone (eventueel veralgemeende) lineaire regressiemodellen aantreft. Informeel kan gezegd worden dat we bij deze effecten aannemen dat ze algemeen geldig zijn, abstractie gemaakt van specifieke instanties zoals bepaalde proefpersonen of bepaalde stimuli. Toevalueffecten, daarentegen, modelleren stimulus- of proefpersoonspecifieke bronnen van variatie die niet onder controle gebracht konden worden. Omdat het een logistisch model betreft, gebeurt het modelleren via een koppelfunctie, nl. de logistische functie. Het model rekent dus met zogenaamde logits, die later desgewenst naar waarschijnlijkheden kunnen worden omgerekend.

Er zijn twee soorten toevalueffecten: *random intercepts* en *random slopes*. Met random intercepts kunnen de baselinewaarden van bijvoorbeeld individuele stimuli of proefpersonen worden bijgesteld. In ons geval, bijvoorbeeld, zou de baselinewaarde van proefpersonen die betere cognaatvertalers zijn dan men op basis van de vaste effecten zou denken enigszins opgeschroefd kunnen worden. Hetzelfde geldt mutatis mutandis voor de stimuli: Doelwoorden die ceteris paribus systematisch beter verstaanbaar zijn, zouden een hogere baselinewaarde kunnen krijgen. Met random slopes kan de variabiliteit van de vaste effecten in kaart worden gebracht. Kennis van het Engels kan bijvoorbeeld in het algemeen bijdragen tot correcte cognaatherkenning, maar de grootte van dit effect kan van doelwoord tot doelwoord verschillend zijn. Bij stimuli waarvoor kennis van het Engels een kleinere rol speelt, brengt het toevalueffect dan een negatieve correctie aan het betreffende vaste effect; bij stimuli waarvoor kennis van het Engels een grotere rol speelt, zal deze correctie positief zijn. De noodzakelijkheid van random intercepts en slopes kan vastgesteld worden door de fit van een model met het toevalueffect te vergelijken met die van een model zonder toevalueffect, bijvoorbeeld met behulp van de AIC-statistiek (Akaike, 1974) of een likelihood ratio-test.

Wanneer er voor de doelwoorden zowel random intercepts als random slopes gemodelleerd worden, bestaat de mogelijkheid dat deze bijstellingen met elkaar gecorreleerd zijn. Een parameter die deze correlatie modelleert kan daarom in het model worden opgenomen. De noodzakelijkheid van deze parameter kan eveneens worden vastgesteld door de fit van een model met en een model zonder deze parameter met elkaar te vergelijken.

Toevalueffecten kunnen hiërarchisch gemodelleerd worden. Wanneer de proefpersonen bijvoorbeeld scholieren zijn, kunnen de toevalueffecten

van de individuen worden ondergebracht bij toevalseffecten voor klassen, die op hun beurt onder toevalseffecten voor scholen kunnen staan. Zo kan systematische variabiliteit die niet met behulp van vaste effecten gemodelleerd werd eerst worden gelinkt aan de school, vervolgens aan de klas en vervolgens aan het individu. Hetzelfde principe geldt voor talige stimuli en de taal waaruit ze zijn ontleend. Het modelleren van toevalseffecten voor de talen zelf kan gerechtvaardigd worden wanneer men bedenkt dat proefpersonen taalspecifieke kennis of voorstellingen, bijvoorbeeld over grafeem-foneemkoppelingen, kunnen toepassen bij het vertalen van cognaten. Als de talen, hiërarchisch gemodelleerd boven de stimuli zelf, een robuust systematisch effect uitoefenen op de baseline of op de grootte van de vaste effecten, dan zal een model met toevalseffecten voor zowel de stimulustalen als de stimuli een betere relatieve fit vertonen dan een model met enkel toevalseffecten voor de stimuli. Door het vergelijken van modellen met en zonder toevalseffecten voor de stimulustalen kunnen we met andere woorden nagaan in hoeverre het effect van predictoren in Lx-cognaatherkenning van Lx tot Lx verschillen.

4. Resultaten

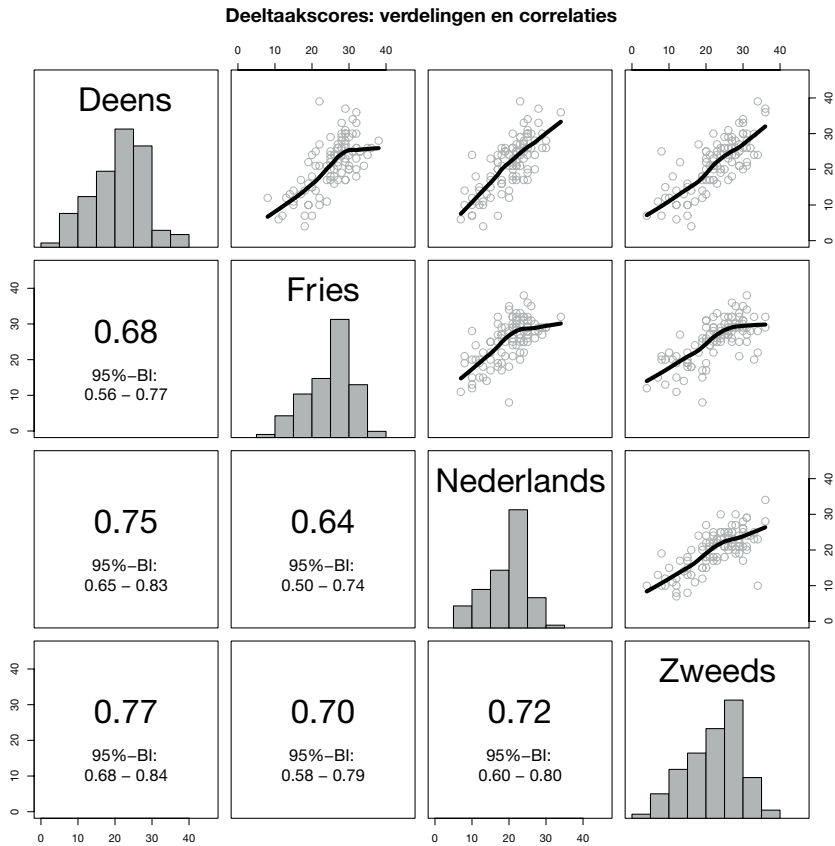
4.1 Individuele scores

Tabel 2 geeft aan hoeveel doelwoorden de proefpersonen per taal correct wisten te vertalen en figuur 2 toont de verdeling van de individuele scores per taal en hun onderlinge correlaties. Deze intercorrelaties zijn substantieel en van ongeveer dezelfde orde ($r \approx 0,7$), wat suggereert dat de vaardigheid om cognaten in een bepaalde Lx te herkennen en te vertalen ten dele afhangt van een latente vaardigheid die onafhankelijk is van de Lx in kwestie. Bij de GLMM-modelleringen zullen we nagaan in hoeverre de effecten van persoonsgerelateerde variabelen door de stimulustaal mee bepaald worden.

Tabel 2 Beschrijvende statistieken van het aantal juist vertaalde doelwoorden per taal per proefpersoon (n = 98)

In het Nederlands waren er 46 cognaten te vertalen, in de andere talen 45.

	Spreidingsbreedte	Mediaan	Gemiddelde	Standaardafwijking
Deens	4–39	22,5	21,5	7,3
Fries	8–38	27,0	25,6	6,0
Nederlands	7–34	21,0	20,1	5,6
Zweeds	4–36	23,5	22,7	7,2



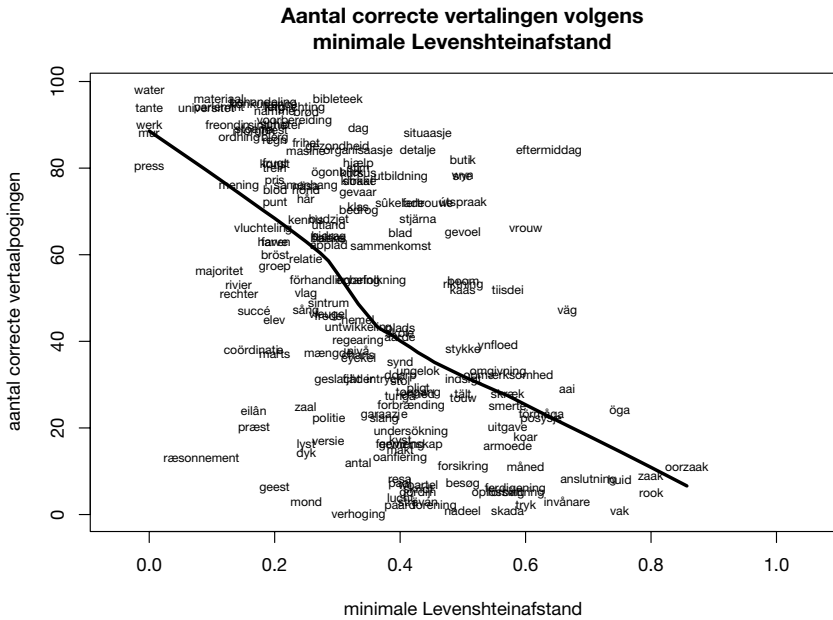
Figuur 2 Verdeling van en correlaties tussen deeltaakscores met hun 95%-betrouwbaarheidsinterval ($n = 98$). De maximumscore voor Deens, Fries en Zweeds was 45, voor Nederlands 46.

4.2 Verkennende bivariate analyses

4.2.1 Orthografische afstand

Het aantal correcte vertalingen per doelwoord ($n = 181$) correleert met de Levenshteinafstand tot hun cognaat in het Duits ($r = -0,42$), in het Engels ($r = -0,39$) en in het Frans ($r = -0,28$). Het is echter de minimale Levenshteinafstand die de sterkste samenhang met het aantal correcte vertalingen vertoont ($r = -0,57$, zie figuur 3). Dit kan uiteraard worden verklaard door het feit dat onze proefpersonen meertalig zijn en bijgevolg ook doelwoorden juist kunnen vertalen wanneer deze geen of geen transparant cognaatver-

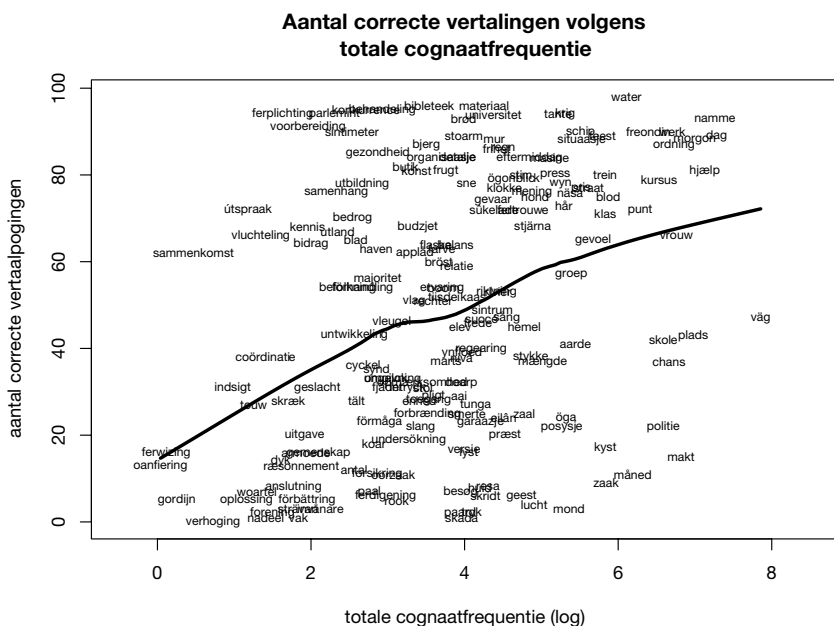
band vertonen met een Duits woord. Voor de uiteindelijke modelleringen beperken we ons daarom tot deze minimale Levenshteinafstand.



Figuur 3 Aantal correcte vertalingen per doelwoord ($n = 181$) in functie van de minimale Levenshteinafstand.

4.2.2 Cognaatfrequentie

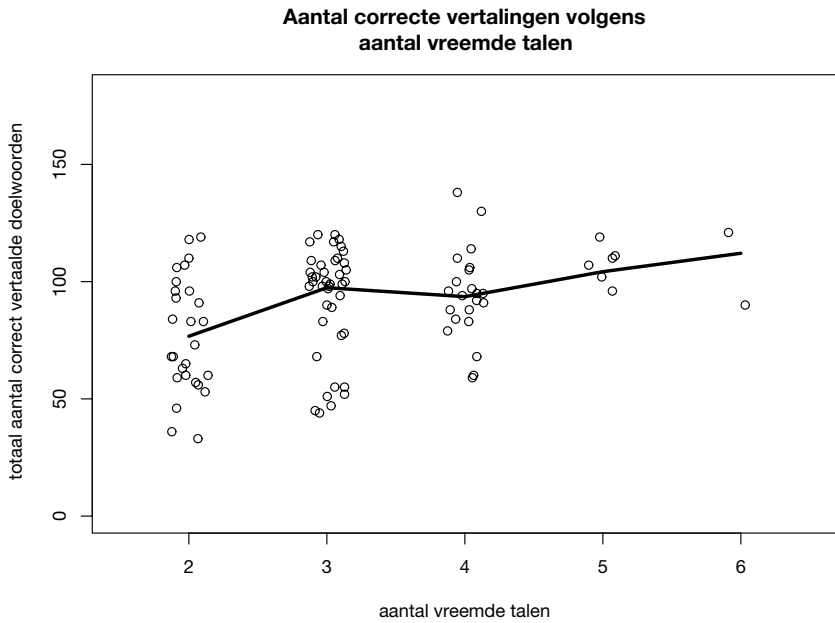
Ook de logaritmisch getransformeerde cognaatfrequentie in het Duits ($r = 0,20$), in het Engels ($r = 0,36$) en in het Frans ($r = 0,23$) correleren met het aantal keer dat een doelwoord juist wordt vertaald. De drietalige maat, de logaritmisch getransformeerde geaggregeerde cognaatfrequentie, vertoont een samenhang met de verstaanbaarheid van de doelwoorden die ongeveer even sterk is als deze van Engelse cognaatfrequentie ($r = 0,34$, zie figuur 4). Daarom opteren we, net als bij de orthografische afstanden, voor deze drietalige maat bij onze modelleringen.



Figuur 4 Aantal correcte vertalingen per doelwoord ($n = 181$) in functie van de totale cogaatfrequentie (logaritmisch getransformeerd)

4.2.3 Aantal beheerste talen

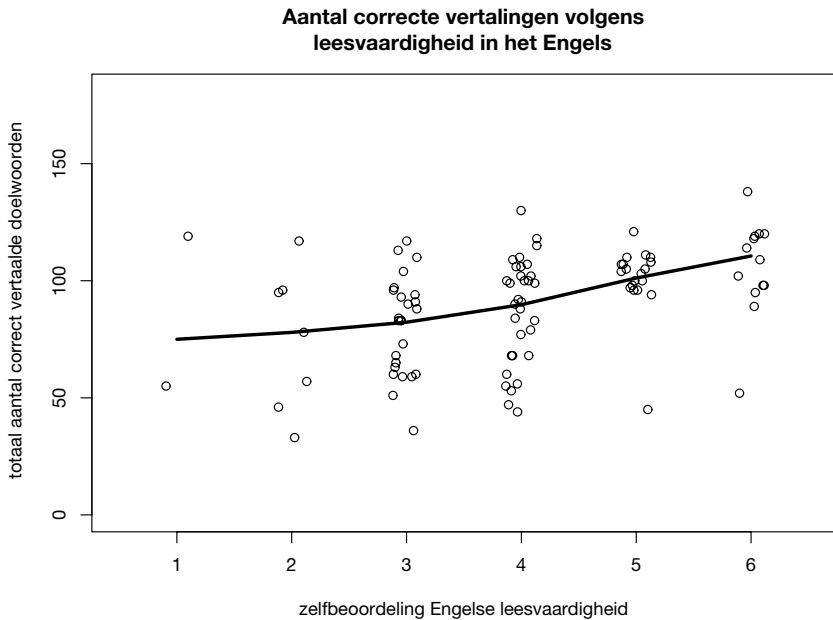
Net als bij Berthele (2011) en Berthele en Lambelet (2009) correleert het aantal vreemde talen dat de proefpersonen ($n = 98$) in meer of mindere mate beheersen met het totale aantal doelwoorden dat ze juist wisten te vertalen ($r = 0,34$, zie figuur 5). Merk op dat het daarbij met uitzondering van het Engels uitsluitend over niet-Germaanse vreemde talen gaat.



Figuur 5 Aantal correcte vertalingen per proefpersoon (n = 98) in functie van het aantal vreemde talen

4.2.4 Kennis van het Engels

Het aantal correct vertaalde doelwoorden per proefpersoon correleert positief met de zelf ingeschatte leesvaardigheid in het Engels ($r = 0,39$, zie figuur 6). De correlatie met de zelf ingeschatte luistervaardigheid in het Engels ligt enigszins lager ($r = 0,27$) en omdat zelf ingeschatte lees- en luistervaardigheden sterk met elkaar gecorreleerd zijn ($r = 0,88$), zullen we bij de modelleringen enkel rekening houden met de leesvaardigheid. De zelf ingeschatte lees- en luistervaardigheid in het Frans correleert overigens amper met het aantal correcte vertalingen ($r = 0,15$ resp. $0,13$) en wordt in ons verslag verder buiten beschouwing gelaten.¹⁰



Figuur 6 Aantal correcte vertalingen per proefpersoon ($n = 98$) in functie van de zelfingeschatte leesvaardigheid in het Engels

4.3 Modelling met GLMM's

4.3.1 Modelling zonder toevalseffecten voor stimulistalen

We modelleerden de correctheid van de 17.738 vertaalpogingen (98 proefpersonen \times 181 doelwoorden per proefpersoon) in functie van gekruiste toevalseffecten (random intercepts) voor de proefpersonen en stimuli en voegden stapsgewijs de vier variabelen die we hierboven hebben besproken (zonder onderlinge interacties) en de relevante random slopes aan het model toe. Bij het toevoegen van de random slopes werden tevens overall correlatieparameters tussen de toevalseffecten gemodelleerd. De variabelen werden om technische redenen eerst gecentreerd rond hun rekenkundig gemiddelde (zie Baayen, 2008: 254-255, voor de logica hierachter). De modellen werden met elkaar vergeleken aan de hand van hun AIC- en log-likelihood-waarden (zie tabellen 3 en 4); modellen met lagere AIC-waarden en hogere log-likelihood-waarden worden hierbij als betere modellen beschouwd.

Tabel 3 De bijdrage van de vaste effecten tot het model beginnend bij een model met gekruiste random intercepts voor de doelwoorden en de proefpersonen
 Alle variabelen werden eerst rond hun rekenkundig gemiddelde gecentreerd.

Toegevoegd vast effect	Afname AIC	Toename log-LR	LR-test (p-waarde)	Commentaar
Random intercepts voor doelwoorden en proefpersonen				
+ minimale Levenshtein-afstand	70	36	< 0,001	
+ totale cognaatfrequentie (log)	15	9	< 0,001	
+ zelfbeoordeling Engelse leesvaardigheid	14	8	< 0,001	
+ aantal vreemde talen	5	4	< 0,01	Alle vaste effecten dragen bij tot het model en worden erin opgenomen.

Tabel 4 De bijdrage van proefpersoon- en doelwoordspecifieke random slopes tot het model beginnend bij een model met gekruiste random intercepts voor de doelwoorden en de proefpersonen en de vaste effecten uit tabel 3
 Alle variabelen werden eerst rond hun rekenkundig gemiddelde gecentreerd.
 Overall werden parameters die de correlaties tussen de toevalseffecten modelleren aan het model toegevoegd.

Toegevoegde random slope	Afname AIC	Toename log-LR	LR-test (p-waarde)	Commentaar
Vaste effecten en random intercepts				
+ minimale Levenshtein-afstand (proefpersoon)	71	37	< 0,001	Impact van de minimale Levenshtein-afstand hangt van de proefpersoon af.
+ zelfbeoordeling Engelse leesvaardigheid (doelwoord)	7	5	< 0,01	Impact van Engelse leesvaardigheid varieert per woord.
+ aantal vreemde talen (doelwoord)	2	4	0,04	Impact van het aantal min of meer beheerste vreemde talen varieert per woord.
+ totale cognaatfrequentie (log) (proefpersoon)	1	2	0,19	Rol van totale cognaatfrequentie lijkt niet van proefpersoon tot proefpersoon te verschillen, dus geen random slope voor totale cognaatfrequentie nodig in model.
- totale cognaatfrequentie (log) (proefpersoon)	-1	-2		Uiteindelijk model.

Uit dit modelvergelijkingsproces blijkt dat het optimale model naast gekruiste random intercepts ook vaste effecten voor de vier besproken variabelen bevatten. Deze vaste effecten worden beschreven in tabel 5 en hun partiële effecten worden weergegeven in figuur 7. De richting van al deze vaste effecten ligt in de lijn van de verwachtingen: negatief voor Levenshteinafstand en positief voor cognaatfrequentie, Engelse leesvaardigheid en aantal vreemde talen. Zowel uit de AIC-verschillen uit tabel 3 als uit de effectgrootten uit tabel 5 en figuur 7 blijkt dat van deze vier variabelen de minimale Levenshteinafstand de grootste impact op correcte cogaatherkenning heeft.

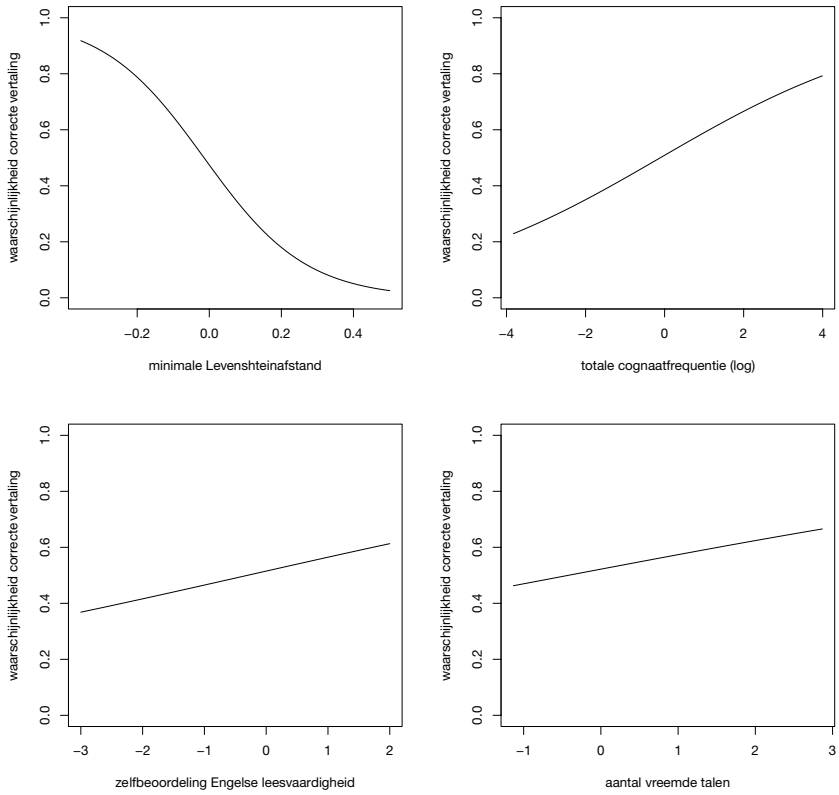
Tabel 5 De vaste effecten in het logistische lineaire model met gemengde effecten.

De variabelen zijn gecentreerd rond hun rekenkundig gemiddelde. De parameters, standaardfouten en effectgrootten worden uitgedrukt in logits. De effectengrootten zijn het absolute verschil (in logits) tussen de waarde van de afhankelijke variabele voor het minimum van de betreffende onafhankelijke variabele en die voor het maximum ervan.

Variabele	Gecentreerd rond	Parameter \pm SE	p -waarde	Effectgrootte
(snijpunt)		$-0,10 \pm 0,15$	0,51	
minimale Levenshteinafstand	0,36	$-7,07 \pm 0,76$	$< 0,001$	6,06
totale cognaatfrequentie (log)	3,86	$0,33 \pm 0,08$	$< 0,001$	2,55
zelfbeoordeling Engelse leesvaardigheid	4,00	$0,20 \pm 0,06$	0,001	1,00
aantal vreemde talen	3,13	$0,20 \pm 0,08$	0,008	0,84

Daarenboven toont tabel 4 dat ook model random slopes voor de minimale Levenshteinafstand, leesvaardigheid in het Engels en het aantal vreemde talen in het model moeten worden opgenomen. Concreet betekent dit dat het precieze effect van de minimale Levenshteinafstand van proefpersoon tot proefpersoon verschilt, net zoals de precieze effecten van leesvaardigheid en grootte van het taalrepertoire van doelwoord tot doelwoord verschillen. Het effect van cognaatfrequentie bleek echter niet sterk genoeg van proefpersoon tot proefpersoon te variëren om er random slopes voor te rechtvaardigen. De modelparameters voor deze toevalseffecten, d.w.z. hun standaardafwijking, worden gegeven in tabel 6.

Partiële vaste effecten



Figuur 7 Partiële vaste effecten voor het optimale logistische model met gemengde effecten (uitgedrukt in waarschijnlijkheden). De variabelen werden rond hun rekenkundig gemiddelde gecentreerd.

Tabel 6 De toevalseffecten in het logistische lineaire model met gemengde effecten. Toevalseffecten komen steeds uit een gaussverdeling met gemiddelde 0 en een bepaalde gemodelleerde standaardafwijking σ (in dit geval uitgedrukt in logits).

Groep	Variabele	σ
doelwoorden	(snijpunt)	1,64
	zelfbeoordeling Engelse leesvaardigheid	0,12
	aantal vreemde talen	0,16
proefpersonen	(snijpunt)	0,85
	minimale Levenshteinafstand	1,64

4.3.2 *Modellering met toevalseffecten voor stimulustalen*

Zoals tabel 7 toont, leidt het hiërarchisch modelleren van de doelwoorden onder hun stimulustaal met behulp van een random intercept niet tot een beter model. De ceteris paribus-verstaanbaarheid van een doelwoord lijkt in deze studie dus niet systematisch af te hangen van de stimulustaal, althans niet sterk genoeg om ervoor een extra parameter aan het model toe te voegen. Belangrijker is echter dat geen van de vier vaste effecten systematisch bijgesteld moet worden naar gelang de stimulustaal met behulp van random slopes. Dit betekent dat de impact van de vier onderzochte variabelen niet substantieel varieert van stimulustaal tot stimulustaal, ook al is er woord- en proefpersoongerelateerde variabiliteit.¹¹

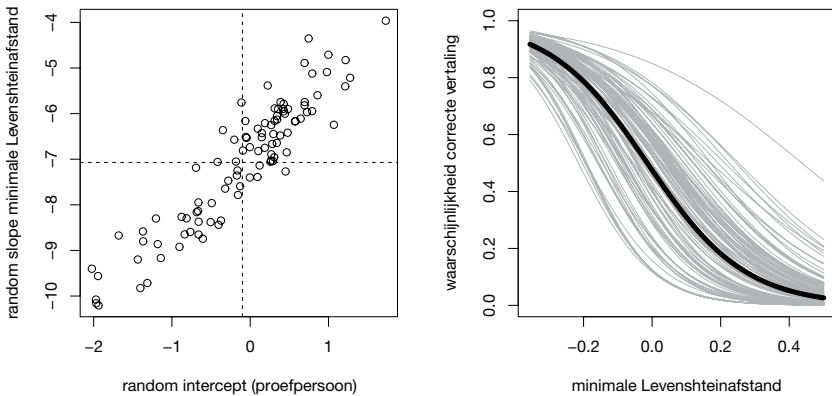
Tabel 7 De bijdrage van stimulustaalspecifieke toevalseffecten tot het model beginnend bij een model met de vaste effecten uit tabel 3 en de proefpersoon- en doelwoordspecifieke toevalseffecten uit tabel 4

Alle variabelen werden eerst rond hun rekenkundig gemiddelde gecentreerd.

Toegevoegd toevals-effect	Afname AIC	Toename log-LR	LR-test (<i>p</i> -waarde)	Commentaar
Vaste effecten en random effecten				
+ random intercept (stimulustaal)	-1	1	0,3	Geen random intercepts voor stimulustalen nodig. Wordt vooralsnog in model behouden omdat random slopes gemodelleerd worden.
+ minimale Levenshtein-afstand (stimulustaal)	-4	0	> 0,99	Rol van minimale Levenshteinafstand lijkt niet van stimulustaal tot stimulustaal te verschillen, dus geen random slopes voor minimale Levenshteinafstand per stimulustaal nodig in model.
- minimale Levenshtein-afstand (stimulustaal)	4	0		
+ totale cognaatfrequentie (log) (stimulustaal)	-4	0	0,97	Rol van totale cognaatfrequentie lijkt niet van stimulustaal tot stimulustaal te verschillen, dus geen random slopes voor totale cognaatfrequentie per stimulustaal nodig in model.
- totale cognaatfrequentie (log) (stimulustaal)	4			
+ zelfbeoordeling Engelse leesvaardigheid (stimulustaal)	0	2	0,16	Rol van Engelse leesvaardigheid lijkt niet van stimulustaal tot stimulustaal te verschillen, dus geen random slopes voor Engelse leesvaardigheid per stimulustaal nodig in model.
- zelfbeoordeling Engelse leesvaardigheid (stimulustaal)	0	-2		
+ aantal vreemde talen (stimulustaal)	-2	1	0,49	Rol van aantal vreemde talen lijkt niet van stimulustaal tot stimulustaal te verschillen, dus geen random slopes voor aantal vreemde talen per stimulustaal nodig in model.
- aantal vreemde talen (stimulustaal)	2	-1		
- random intercept (stimulustaal)	1	-1		Uiteindelijk model.

4.3.3 Excursie: Intercorrelaties tussen toevalseffecten

Interessant genoeg bevat het model ook een parameter die een substantiële correlatie modelleert tussen de proefpersoonspecifieke bijstellingen aan het snijpunt en die aan het effect voor de minimale Levenshteinafstand ($\rho = 0,82$; zie figuur 8 links).¹² Wordt deze correlatieparameter uit het model verwijderd, dan stijgt (verslechtert) de AIC-waarde met 37. Ook een likelihood ratio-test geeft aan dat de correlatieparameter gerechtvaardigd is ($\chi^2(1) = 39,2$; $p < 0,001$). Proefpersonen die ceteris paribus meer kans hebben om een willekeurig doelwoord correct te vertalen (en bijgevolg als betere cognaaitherkenners kunnen worden beschouwd, althans ceteris paribus), ondervinden dus in de regel ook een minder negatief effect van een grotere orthografische afstand tussen het doelwoord en een cognaat ervan (zie figuur 8 rechts).



Figuur 8 Links: Verband tussen de proefpersoonspecifieke snijpunten en het proefpersoonspecifieke effect van de minimale Levenshteinafstand ($\rho = 0,82$; uitgedrukt in logits). De stippellijnen geven de waarden van de vaste effecten aan. **Rechts:** Interindividuele variabiliteit van het effect van de minimale Levenshteinafstand (uitgedrukt in waarschijnlijkheden). Elke dunne lijn representeert de regressiecurve van een specifieke proefpersoon. De dikke curve geeft het vaste effect weer. De Levenshteinvariabele is gecentreerd rond haar rekenkundig gemiddelde (0,36).

Ook voor het verband tussen doelwoordspecifieke bijstellingen aan het snijpunt enerzijds en de effecten van leesvaardigheid in het Engels ($\rho = 0,33$) en het aantal vreemde talen ($\rho = 0,20$) werden positieve, maar beduidend lagere, correlatieparameters gemodelleerd. Het verwijderen van deze correlatie-

parameters leidde niet tot een slechtere fit van het model: de AIC-waarde daalt (verbetert) zelfs met 2. Een likelihood ratio-test geeft bovendien geen significant verschil tussen de fit van het complexere model met de extra parameters en die van het eenvoudigere aan ($\chi^2(3) = 4,0; p = 0,27$). We kiezen bijgevolg het eenvoudigere model zonder de doelwoordspecifieke correlatieparameters. Hierdoor wijzigen de parameters van de vaste en toevalseffecten nauwelijks.

5. Bespreking en conclusies

5.1 Beïnvloedende factoren

5.1.1 *Orthografische afstand*

Net als in verschillende eerdere studies met geschreven of gesproken stimuli (Berthele, 2011; Berthele en Lambelet, 2009; Doetjes en Gooskens, 2009; Gooskens et al., 2011; Kürschner et al., 2008; Van Bezooijen en Gooskens, 2005a) was een maat van de formele overlap tussen het Lx-woord en een verwant en gekend vertalingsequivalent een respectabele predictor van de verstaanbaarheid van Lx-woorden: Hoe groter de formele (orthografische) overlap, hoe vaker het Lx-woord correct vertaald wordt. Vergeleken met de andere onderzochte factoren was deze maat van formele overlap ook veruit de belangrijkste beïnvloedende factor, zoals blijkt uit de verschillen in AIC-waarden (tabel 3) en de effectgrootten (tabel 5). Dit rijmt met de bevinding van Kürschner et al. (2008) die de formele (fonetische) overlap tussen Zweedse woorden en hun Deense cognaten als de belangrijkste stimulusgerelateerde voorspeller van de verstaanbaarheid van gesproken Zweedse woorden voor Denen identificeerden. De noodzakelijkheid van random slopes voor de Levenshteinvariabele toont echter aan dat de impact van formele discrepantie substantieel van de proefpersoon in kwestie afhangt.

5.1.2 *Cognaatfrequentie*

Ook de geaggregeerde relatieve frequentie van verwante vertaalequivalenten in drie talen die al onze proefpersonen tot op zekere hoogte beheersten (Duits, Engels en Frans) speelde een rol. Lx-woorden met hoogfrequente cognaten worden vaker correct vertaald dan Lx-woorden met minder frequente cognaten. Voor cognaatfrequentie was er ook een bescheiden

rol weggelegd in de studie van Kürschner et al. (2008), maar voor zover we weten is dit de eerste keer dat de frequentiefactor bij de herkenbaarheid van geschreven doelwoorden werd onderzocht.

5.1.3 *Vreemdetalenkennis*

In lijn met de verwachtingen werd cognaatherkenning in niet beheerste Germaanse talen mee beïnvloed door het taalbeheersingsniveau in een Germaanse vreemde taal, in casu het Engels. Net als Berthele (2011) werkten we hier met zelfbeoordelingen, maar ook Vanhove en Berthele (aangenomen) vonden dat kennis van het Engels, deze keer gemeten met een test, cognaatherkenning in een Germaanse Lx ten goede komt en dat de grootte van het effect niet systematisch afhangt van de modaliteit waarin de doelwoorden worden gepresenteerd.

Net als Vanhove en Berthele (aangenomen) vonden we dat het precieze effect van kennis van het Engels afhangt van het doelwoord in kwestie. Welke stimulusgerelateerde factoren het precieze effect mee bepalen, weten we vooralsnog niet: er is in elk geval geen sterk verband tussen de random slope-waarde enerzijds en de Levenshteinafstand tussen het doelwoord en zijn Engels cognaat ($r_s = -0,02$) of de frequentie van het Engelse cognaat ($r_s = -0,07$) anderzijds.

Tot slot levert deze studie indicaties op dat cognaatherkenning door de band genomen positief wordt beïnvloed door de algemene talenkennis van de proefpersonen, ook wanneer de proefpersonen, afgezien van het Engels, geen verdere vreemde talen beheersen die nauw met de Lx verwant zijn. Positieve correlaties tussen de grootte van het taalrepertoire en de vaardigheid Lx-woorden te vertalen werden eerder al vastgesteld door Berthele en Lambelet (2009) en Berthele (2011), maar in deze studies werden proefpersonen met kennis van vreemde talen die nauw met de Lx'en in kwestie (Roemeens en Reto-Romaans resp. Deens en Zweeds) niet systematisch van het onderzoek uitgesloten. Onze studie suggereert dus iets sterker dat niet alleen kennis van verwante vreemde talen maar ook van vreemde talen in het algemeen een bijdrage levert tot de cognaatherkenningsvaardigheid. Dat gezegd zijnde varieert ook de precieze sterke van dit effect al naar gelang het specifieke doelwoord, zonder dat we hiervoor al een verantwoordelijke stimulusgerelateerde factor konden identificeren.

5.2 **Lx-specifieke invloeden**

In onze analyses vonden we niets dat erop wees dat de Lx waaruit een doelwoord werd ontleend een systematische invloed uitoefent op het effect van de vier onderzochte variabelen. Dit duidt erop dat bevindingen uit

studies naar Lx-cognaatherkenning een groter geldingsbereik hebben dan de specifieke Lx die onderzocht werd, op zijn minst voor de geschreven modaliteit.

Hierbij dienen we echter te vermelden dat er tussen Duitstalige Zwitsers enerzijds en Deens-, Fries-, Nederlands- en Zweedstaligen anderzijds geen traditie bestaat om met elkaar in de respectievelijke L1 te communiceren. Dit contrasteert met de situatie in bijvoorbeeld Scandinavië, waar Denen, Noren en Zweden vaak (en met variërend succes) hun moedertaal hanteren in zowel gesproken als in geschreven onderlinge communicatie (Bø, 1978; Delsing en Lundin Åkesson, 2005; Haugen, 1966; Maurud, 1976). Een dergelijke traditie van receptieve meertaligheid kan resulteren in een betere vertrouwdeheid met de relevante correspondentieregels tussen de L1 en de Lx (Bannert, 1981). Dit zou in het bijzonder het effect van de formele overlap danig kunnen verkleinen. Op die manier zouden er Lx-specifieke effecten kunnen ontstaan. Onze resultaten kunnen bijgevolg niet zonder meer worden gegeneraliseerd naar dergelijke constellaties.

5.3 Eentalige versus meertalige predictoren

In deze studie hebben we gebruik gemaakt van Levenshteinafstanden en cognaatfrequenties die op meerdere talen, namelijk Duits, Frans en Engels, gebaseerd zijn omdat bleek dat deze ietwat sterker correleren met het aantal correcte vertaalpogingen per doelwoord dan predictoren die enkel op het Duits gebaseerd zijn. Dit geldt zelfs wanneer we ons beperken tot doelwoorden die een Duits cognaat hebben ($n = 176$): de Levenshteinafstand tot het Duits correleert weliswaar relatief sterk met het aantal juiste vertalingen per doelwoord ($r = -0,51$), maar moet de duimen leggen voor de minimale Levenshteinafstand ($r = -0,57$; intercorrelatie tussen Duitse en minimale Levenshteinafstand: $r = 0,86$) en ook de correlatie met Duitse cognaatfrequentie ($r = 0,23$) is enigszins zwakker dan die met de totale cognaatfrequentie ($r = 0,33$; intercorrelatie tussen Duitse en totale cognaatfrequentie: $r = 0,85$). De GLMM voor deze 176 doelwoorden vertoont bovendien een betere fit wanneer het de meertalige predictoren bevat dan wanneer het enkel de Duitse predictoren bevat.¹³

Dat proefpersonen zich bij Lx-cognaatherkenning niet op één enkele taalvariëteit beroepen, is op zich geen nieuw gegeven (zie bv. Berthele, 2011; Gooskens et al., 2011; Kürschner et al., 2008). Vreemdetalenkennis wordt echter lang niet altijd systematisch in acht genomen bij onderzoek naar cognaaatherkenning. Op basis van onze studie willen we daarom een lans breken voor de inachtneming van zulke kennis en voor de verdere verfijning van variabelen die uit meerdere talen zijn afgeleid. Wat dit laatste betreft

zou bijvoorbeeld de belangrijkheid van de brontaal bij het berekenen van een globale cognaatfrequentiegraad kunnen worden volgens het taalbeheersingsniveau van de proefpersonen of haar psychotypologische afstand (Kellerman, 1983) tot de Lx.

5.4 Duiding van intercorrelaties tussen toevalseffecten

Tot slot nemen we de gemodelleerde intercorrelaties tussen de random intercepts en de random slopes onder de loep. De intercorrelaties tussen de doelwoordspecifieke bijstellingen waren eerder laag en rechtvaardigden geen correlatieparameters in het model. Er is bijgevolg nog geen sprake van systematische covariatie die verklaard moet worden. Proefpersonen met positieve bijstellingen aan hun individueel snijpunt en die dus ceteris paribus goede cognaatvertalers zijn, bleken echter positieve bijstellingen aan het richtingscoëfficiënt voor het effect van de Levenshteinafstand te hebben ($\rho = 0,82$). Het vaste effect van de Levenshteinafstand is negatief, dus betekent dit dat ze in de regel minder hinder ondervinden van grotere Levenshteinafstanden, zoals afgebeeld in figuur 8 (rechts). We willen erop wijzen dat dit verband geen wiskundige noodzakelijkheid is. Het is namelijk perfect plausibel dat betere cognaattherkenning zich enkel uit in een hogere basiswaarschijnlijkheid op een correcte vertaling, maar verder onafhankelijk is van de gevoeligheid voor orthografische overlap.

Deze bevinding strookt met Bertheles (2008, 2011) idee van de *Wahrnehmungstoleranz*: wie toleranter tegenover formele discrepanties tussen Lx-woorden en hun L1-, L2-, ..., Ln-cognaten is, kan in de regel beter cognaten herkennen, op voorwaarde dat implausibele vertaalkandidaten verworpen kunnen worden (*Fokussierungsfähigkeit*). Een interessante piste voor verder onderzoek is de factoren te identificeren die deze *Wahrnehmungstoleranz* en *Fokussierungsfähigkeit* bepalen.

Noten

1. Corresponderende auteur: Jan Vanhove, Studienbereich Mehrsprachigkeitsforschung und Fremdsprachendidaktik, Universität Freiburg, Rue de Rome 1, CH-1700 Fribourg, Zwitserland. Telefoon: +41 (0) 26 300 7141. E-mail: jan.vanhove@unifr.ch.
We willen Charlotte Gooskens (Afdeling Toegepaste taalwetenschap, Rijksuniversiteit Groningen) hartelijk bedanken voor het ter beschikking stellen van een meertalige lijst met frequente substantieven en hun Germaanse vertalingen. We bedanken ook twee anonieme lezers van T&T voor hun nuttige commentaren.
Deze studie werd ontworpen in het kader van het werkcollege *Plurilinguale Kompetenz* (Universiteit van Fribourg, voorjaar 2012). De data werden verzameld door de studenten, die we bij deze dan ook van harte willen bedanken (in willekeurige volgorde): Christof

Chesini, Annina Keller, Melahat Yapici, Nathalie Bagnoud, Barbara Grüter, Carlos Pestana, Thomas Michel, Magalie Desgrippes, Fabienne Strässle, Marylin Krieg, Barbara Rampolla en Katharina Karges.

De data en computercode die voor de analyses gebruikt werden, zijn beschikbaar op <http://dx.doi.org/10.6084/m9.figshare.1063422>.

2. We willen er op wijzen dat lang niet alle auteurs ook zulke generalisaties wilden maken. Kürschner, Gooskens en Van Bezooijen (2008) en Doetjes en Gooskens (2009), bijvoorbeeld, zijn specifiek geïnteresseerd in de herkenning van Zweedse cognaten door Denen.
3. Zie Heeringa (2004: 130–132) voor andere normalisatieprocedures.
4. Het Levenshteinalgoritme kan verfijnd worden door bijvoorbeeld een substitutie van *t* door *d* minder zwaar te laten doorwegen dan een substitutie van *t* door *w* (naar het voorbeeld van Beijering, Gooskens en Heeringa, 2008, en Gooskens et al., 2008), maar bij het modelleren van cognatherkenning voert dit zelden tot substantiële verbeteringen, althans niet in de geschreven modaliteit (aldus ook Möller, 2012).
5. <http://europass.cedefop.europa.eu/de/resources/european-language-levels-cefr>
6. Drie proefpersonen vermeldden geen dialectregio. Vier proefpersonen gaven twee dialectregio's op.
7. Het idee was telkens vijf profielwoorden per taal te selecteren, maar we werden er ons pas later van bewust dat het Nederlandse woord *touw*, dat normaal gezien in het Duits als *Seil* vertaald wordt, ook als *Tau* kan worden vertaald. Bijgevolg was het geen geschikt profielwoord, maar een doelwoord.
8. Beide databases zijn gratis te downloaden op <http://crr.ugent.be/programs-data/subtitle-frequencies/>.
9. Lexique 3 is gratis te downloaden op <http://www.lexique.org/>.
10. Wanneer leesvaardigheid in het Frans toch in de modellen wordt opgenomen, leidt dit niet tot betere modellen. Het effect van Franse leesvaardigheid is m.a.w. niet significant. De parameters van de modellen veranderen bovendien nauwelijks wanneer Franse leesvaardigheid in de modellen wordt opgenomen.
11. Met slechts vier verschillende waarden is de variabele 'stimulustaal' mogelijk niet divers genoeg voor een zinvolle modellering van toevalseffecten. Een alternatieve oplossing bestaat erin stimulustaal als vast effect te modelleren en te laten interageren met de andere vaste effecten. Conceptueel is dit voor onze doeleinden minder aantrekkelijk omdat hiermee de indruk wordt gewekt dat onze vier stimulustalen de enige mogelijke stimulustalen zijn. Aan onze bevindingen verandert het modelleren van stimulustaal als vast effect echter niets, zoals geïnteresseerde lezers kunnen nagaan aan de hand van de data en computercode die beide online beschikbaar zijn.
12. Deze ρ -coëfficiënt is een modelparameter. De (steekproef)correlatie tussen de daadwerkelijke bijstellingen, zoals weergegeven in figuur 8, wordt weliswaar hiervan afgeleid, maar is er niet aan gelijk (zie Baayen, 2008: 247, voor meer over verschil tussen modelparameters en daadwerkelijke bijstellingen in modellen met gemengde effecten).
13. GLMM's met gekruiste random intercepts voor de doelwoorden en de proefpersonen met vaste effecten voor (gecentreerde) Levenshteinafstand en cognaatfrequentie (log) met een random slope voor Levenshteinafstand per proefpersoon: AIC voor de eenalige variabelen: 15.267; AIC voor de meertalige variabelen: 15.241.
14. Dit is een schrijffout van onzentwege. De correcte Zweedse schrijfwijze is *cykel*.

Bibliografie

- Akaike, Hirotugu (1974). A new look at the statistical model identification. In: *IEEE Transactions on Automatic Control* AC-19, 716-723.
- Baayen, R. Harald (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, Cambridge University Press.
- Baayen, R. Harald, Douglas J. Davidson en Douglas M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. In: *Journal of Memory and Language* 59, 390-412.
- Bannert, Robert (1981). Referat av diskussionen i sektionen Talperceptionsforskning och nordisk hörförståelse. In: Claes-Christian Elert (red.). *Internordisk språkförståelse. Föredrag och diskussioner vid ett symposium på Rungstegegaard utanför Köpenhamn den 24-26 mars 1980*. Umeå, Universitetet i Umeå, p. 37-45.
- Bates, Douglas M., Martin Maechler en Ben Bolker (2012). *lme4: Linear mixed-effect models using Eigen and Eigen*. R-bibliotheek, versie 0.999999-0. Beschikbaar op <http://cran.r-project.org/package=lme4>.
- Beijering, Karin, Charlotte Gooskens en Wilbert Heeringa (2008). Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm. In: *Linguistics in the Netherlands* 25, 13-24.
- Berthele, Raphael (2008). Dialekt-Standard Situationen als embryonale Mehrsprachigkeit. Erkenntnisse zum interlingualen Potenzial des Provinzlerdaseins. In: *Sociolinguistica* 22, 87-107.
- Berthele, Raphael (2011). On abduction in receptive multilingualism: Evidence from cognate guessing tasks. In: *Applied Linguistics Review* 2, 191-219.
- Berthele, Raphael en Amelia Lambelet (2009). Approche empirique de l'intercompréhension : répertoires, processus et résultats. In: *Lidil* 39, 151-162.
- Bø, Inge (1978). *Ungdom og naboland: En undersøkelse av skolens og fjernsynets betydning for nabospråkförståelsen*. Stavanger, Rogalandforskning.
- Bravo, Marco A., Elfrieda H. Hiebert en P. David Pearson (2007). Tapping the linguistic resources of Spanish/English bilinguals: The role of cognates in science. In: R.K. Wagner, A. Muse en K. Tannenman (red.). *Vocabulary development and its implications for reading comprehension*. New York, Guilford, p. 140-156.
- Brybaert, Marc, Matthias Buchmeier, Markus Conrad, Arthur M. Jacobs, Jens Bölte en Andrea Böhl (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. In: *Experimental Psychology* 58, 412-424.
- Brybaert, Marc en Boris New (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. In: *Behavior Research Methods* 41, 977-990.
- Clark, Herbert H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. In: *Journal of Verbal Learning and Verbal Behavior* 12, 335-359.
- Coleman, Edmund B. (1964). Generalizing to a language population. In: *Psychological Reports* 14, 219-226.
- Costa, Albert, Alfonso Caramazza en Nuria Sebastián-Gallés (2000). The cognate facilitation effect: Implications for models of lexical access. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 1283-1296.
- De Bot, Kees (1992). A bilingual production model: Levelt's 'Speaking' model adapted. In: *Applied Linguistics* 13, 1-24.

- Delsing, Lars-Olof en Katarina Lundin Åkesson (2005). *Hållerspråket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska*. Kopenhagen, Nordiska ministerrådet.
- De Groot, Annette M.B. en Rineke Keijzer (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. In: *Language Learning* 50, 1-56.
- Doetjes, Gerald en Charlotte Gooskens (2009). Skriftsprogets rolle i den dansk-svenske talesprogsforståelse. In: *Språk och stil* 19, 105-123.
- Dyen, Isidore, Joseph B. Kruskal en Paul Black (1992). An Indoeuropean classification: A lexicostatistical experiment. In: *Transactions of the American Philosophical Society* 82, 1-132.
- Gooskens, Charlotte (2007). The contribution of linguistic factors to the intelligibility of closely related languages. In: *Journal of Multilingual and Multicultural Development* 28, 445-467.
- Gooskens, Charlotte, Wilbert Heeringa en Karin Beijering (2008). Phonetic and lexical predictors of intelligibility. In: *International Journal of Humanities and Arts Computing* 2, 63-81.
- Gooskens, Charlotte, Sebastian Kürschner en Renée Van Bezooijen (2011). Intelligibility of Standard German and Low German to speakers of Dutch. In: *Dialectologia* Special issue, II, 35-63.
- Gooskens, Charlotte en Renée Van Bezooijen (2006). Mutual comprehensibility of written Afrikaans and Dutch: Symmetrical or asymmetrical? In: *Literary and Linguistic Computing* 21, 543-557.
- Gray, Russell D. en Quentin D. Atkinson (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. In: *Nature* 426, 435-439.
- Grosjean, François (2010). *Bilingual: Life and reality*. Cambridge, Harvard University Press.
- Haugen, Einar (1966). Semicommunication: The language gap in Scandinavia. In: *Sociological Inquiry* 36, 280-297.
- Heeringa, Wilbert (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. Groningen, Groningen Dissertations in Linguistics (Grodil).
- Jaeger, T. Florian (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. In: *Journal of Memory and Language* 59, 434-446.
- Jaeger, T. Florian, Peter Graff, William Croft en Daniel Pontillo (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. In: *Linguistic Typology* 15, 281-320.
- Jensen, John B. (1989). On the mutual intelligibility of Spanish and Portuguese. In: *Hispania* 72, 848-852.
- Kellerman, Eric (1983). Now you see it, now you don't. In: Susan M. Gass en Larry Selinker (red.). *Language transfer in language learning*. Rowley (MA), Newbury House, p. 140-156.
- Kürschner, Sebastian en Charlotte Gooskens (2008). Interskandinavisk ordförståelse: En internetbaseret undersøgelsemetode. In: Marianne Nordman et al. (red.). *Svenskans beskrivning 29. Förhandlingar vid Tjugonionde sammankomsten för svenskans beskrivning. Vasa den 4 och 5 maj 2007*. Vaasa, Svenska-österbottniska samfundet, p. 170-178.
- Kürschner, Sebastian en Charlotte Gooskens (2011). Verstehen nah verwandter Verietäten über Staatsgrenzen hinweg. In: Elvira Glaser et al. (red.). *Dynamik des Dialekts: Wandel und Variation. Akten des 3. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD) (= Zeitschrift für Dialektologie und Linguistik – Beihefte 144)*, p. 161-185.
- Kürschner, Sebastian, Charlotte Gooskens en Renée Van Bezooijen (2008). Linguistic determinants of the intelligibility of Swedish words among Danes. In: *International Journal of Humanities and Arts Computing* 2, 83-100.
- Lemhöfer, Kristin, Ton Dijkstra, Herbert Schriefers, R. Harald Baayen, Jonathan Grainger en Pienie Zwitserlood (2008). Native language influences on word recognition in a second

- language: A megastudy. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34, 12-31.
- Levenshtein, Vladimir Iosifovich (1966). Binary codes capable of correcting deletions, insertions, and reversals. In: *Soviet Physics Doklady* 10, 707-710.
- Lotto, Lorella en Annette M.B. De Groot (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. In: *Language Learning* 48, 31-69.
- Lublinter, Shira en Elfrieda H. Hiebert (2011). An analysis of English-Spanish cognates as a source of general academic language. In: *Bilingual Research Journal* 34, 76-93.
- Luce, Paul A. en David B. Pisoni (1998). Recognizing spoken words: The Neighborhood Activation Model. In: *Ear & Hearing* 19, 1-36.
- Maurud, Øivind (1976). *Nabospråksforståelse i Skandinavia: En undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige*. Stockholm, Nordiska rådet.
- Möller, Robert (2011). Wann sind Kognaten erkennbar? Ähnlichkeit und synchrone Transparenz von Kognatenbeziehungen in der germanischen Interkomprehension. In: *Linguistik online* 46, 79-101.
- Möller, Robert (2012). Germanic sound correspondences in didactic presentation and in reality. Presentatie op *Sociolinguistics Symposium 19*, Freie Universität Berlin, 21-24 augustus.
- Möller, Robert en Ludger Zeevaert (2010). "Da denke ich spontan an Tafel!": Zur Worterkennung in verwandten germanischen Sprachen. In: *Zeitschrift für Fremdsprachenforschung* 21, 217-248.
- New, Boris, Marc Brysbaert, Jean Veronis en Christophe Pallier (2007). The use of film subtitles to estimate word frequencies. In: *Applied Psycholinguistics* 28, 661-677.
- Norris, Dennis (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. In: *Psychological Review* 113, 327-357.
- Norris, Dennis en James M. McQueen (2008). Shortlist B: A Bayesian model of continuous speech recognition. In: *Psychological Review* 115, 357-395.
- Orioles, Vincenzo (2004). Plurilinguisme : modèles interprétatifs, terminologie et retombées institutionnelles. In: *Revue française de linguistique appliquée* 9, 11-30.
- Peirce, Charles Sanders (1931). *The collected papers Vol. V: Pragmatism and pragmaticism*. Cambridge (MA), Harvard University Press.
- R Core Team (2012). *R: A language and environment for statistical computing*. Software, versie 2.15.2. Beschikbaar op <http://www.r-project.org/>.
- Schielzeth, Holger en Wolfgang Forstmeier (2009). Conclusions beyond support: Overconfident estimates in mixed models. In: *Behavioral Ecology* 20, 416-420.
- Schüppert, Anja en Charlotte Gooskens (2011). Investigating the role of language attitudes for perception abilities using reaction time. In: *Dialectologia* Special issue, II, 119-140.
- Simpson, Edward H. (1951). The interpretation of interaction in contingency tables. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 13, 238-241.
- Stegmann, Tilbert D. en Horst G. Klein (1999). *EuroComRom – Die sieben Siebe. Romanische Sprachen sofort lesen können*. Aken, Shaker.
- Van Assche, Eva, Wouter Duyck, Robert J. Hartsuiker en Kevin Diependaele (2009). Does bilingualism change native-language reading? In: *Psychological Science* 20, 923-927.
- Van Bezooijen, Renée en Charlotte Gooskens (2005a). How easy is it for speakers of Dutch to understand Frisian and Afrikaans, and why? In: *Linguistics in the Netherlands* 22, 13-24.
- Van Bezooijen, Renée en Charlotte Gooskens (2005b). Intertalig tekstbegrip. De begripelijkheid van Friese en Afrikaanse teksten voor Nederlandse lezers. In: *Nederlandse Taalkunde* 10, 129-152.
- Van Bezooijen, Renée, Charlotte Gooskens en Sebastian Kürschner (2012). Deens is makkelijker voor Friezen dan voor Nederlanders – feit of fabel? In: Piter Boersma, Goffe T. Jensma en Reinier Salverda (red.). *Philologia Frisica anno 2008. Lêzings fan it achttjinde Frysk Filolo-*

- gekongres fan de Fryske Akademy op 10, 11 en 12 desimber 2008*. Leeuwarden, Afûk/Fryske Akademy, p. 286-298.
- Van Hell, Janet G. en Ton Dijkstra (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. In: *Psychonomic Bulletin & Review* 9, 780-789.
- Van Heuven, Vincent J. (2008). Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. In: *International Journal of Humanities and Arts Computing* 2, 39-62.
- Vanhove, Jan en Raphael Berthele (aangenomen). The lifespan development of cognate guessing skills in an unknown related language. In: *International Review of Applied Linguistics in Language Teaching*.
- Wandruszka, Mario (1981). *Die Mehrsprachigkeit des Menschen*. München, Deutscher Taschenbuch-Verlag.
- Wolff, Hans (1959). Intelligibility and inter-ethnic attitudes. In: *Anthropological Linguistics* 1, 34-41.

Over de auteurs

Jan Vanhove, Studienbereich Mehrsprachigkeitsforschung und Fremdsprachendidaktik, Universität Freiburg, Rue de Rome 1, CH-1700 Fribourg, Zwitserland. Telefoon: +41 (0) 26 300 7141. E-mail: jan.vanhove@unifr.ch.

Raphael Berthele, Université de Fribourg/Universität Freiburg, Zwitserland

Appendix: Stimuli

Stimuli met hun Duitse modelvertalingen en de Duitse, Engelse en Franse cognaten die gebruikt werden bij het berekenen van de Levenshteinafstanden en cognaatfrequenties.

Taal	Stimulus	Duitse vertaling	Duits cognaat	Engels cognaat	Frans cognaat
DK	antal	Anzahl	Anzahl		
	befolkning	Bevölkerung	Bevölkerung		
	besøg	Besuch	Besuch		
	bidrag	Beitrag	Beitrag		
	bjerg	Berg	Berg		
	brød	Brot	Brot	bread	
	butik	Boutique	Boutique	boutique	boutique
	dag	Tag	Tag	day	
	detalje	Detail	Detail	detail	détail
	enhed	Einheit	Einheit		
	farve	Farbe	Farbe		
	flaske	Flasche	Flasche		
	forbrænding	Verbrennung	Verbrennung	burning	
	forening	Verein; Vereinigung	Vereinigung		
	forsikring	Versicherung	Versicherung		
	frugt	Frucht	Frucht	fruit	fruit
	hjælp	Hilfe	Hilfe	help	
	indsigt	Einsicht	Einsicht		
	klokke	Glocke; Uhr	Glocke	clock	cloche
	konkurrence	Konkurrenz	Konkurrenz		concurrency
	kursus	Kurs	Kurs	course	cours
	kyst	Küste	Küste	coast	côte
	lyst	Lust	Lust		
	mængde	Menge	Menge		
	majoritet	Mehrheit	Majorität	majority	majorité
	måned	Monat	Monat	month	mois
	marts	März	März	march	mars
	mening	Meinung; Bedeutung; Sinn	Meinung	meaning	
	opmærksomhed	Aufmerksamkeit	Aufmerksamkeit		
	plads	Platz	Platz	place	place
	pligt	Pflicht	Pflicht		
	pris	Preis	Preis	price	prix
	præst	Priester	Priester	priest	prêtre

Taal	Stimulus	Duitse vertaling	Duits cognaat	Engels cognaat	Frans cognaat
	regn	Regen	Regen	rain	
	ræsonnement	Darlegung	Ræsonnement	reasoning	raisonnement
	sammenkomst	Zusammenkunft	Zusammenkunft		
	skole	Schule	Schule	school	école
	skridt	Schritt	Schritt		
	skræk	Schreck	Schreck		
	smerte	Schmerz	Schmerz		
	sne	Schnee	Schnee	snow	
	stol	Stuhl	Stuhl		
	stykke	Stück	Stück		
	synd	Sünde	Sünde	sin	
	tryk	Druck	Druck		
	kylling	Huhn	(profielwoord)		
	penge	Geld	(profielwoord)		
	seng	Bett	(profielwoord)		
	skov	Wald	(profielwoord)		
	æske	Dose	(profielwoord)		
NL	aarde	Erde	Erde	earth	
	armoede	Armut	Armut		
	bedrog	Betrug	Betrug		
	boom	Baum	Baum		
	coördinatie	Koordination	Koordination	coordination	coordination
	ervaring	Erfahrung	Erfahrung		
	geest	Geist	Geist	ghost	
	gevaar	Gefahr	Gefahr		
	gevoel	Gefühl	Gefühl	feeling	
	gezondheid	Gesundheit	Gesundheit		
	gardijn	Gardine	Gardine		
	hemel	Himmel	Himmel		
	hond	Hund	Hund		
	huid	Haut	Haut		
	kaas	Käse	Käse	cheese	
	kennis	Kenntnis	Kenntnis		
	lucht	Luft	Luft		
	mond	Mund	Mund	mouth	
	nadeel	Nachteil	Nachteil		
	oorzaak	Ursache	Ursache		
	paal	Pfahl	Pfahl	pole	
	paard	Pferd	Pferd		
	politie	Polizei	Polizei	police	police
	rechter	Richter	Richter		

FACTOREN BIJ HET HERKENNEN VAN COGNATEN IN ONBEKENDE TALEN

Taal	Stimulus	Duitse vertaling	Duits cognaat	Engels cognaat	Frans cognaat
	relatie	Verhältnis	Relation	relation	relation
	rook	Rauch	Rauch		
	samenhang	Zusammenhang	Zusammenhang		
	schip	Schiff	Schiff	ship	
	straat	Strasse	Strasse	street	
	tante	Tante	Tante	aunt	tante
	toegang	Zugang	Zugang		
	touw	Seil, Tau	Tau		
	trein	Zug		train	train
	uitgave	Ausgabe	Ausgabe		
	vak	Fach	Fach		
	verhoging	Erhöhung	Erhöhung		
	versie	Version	Version	version	version
	vlag	Flagge	Flagge	flag	
	vleugel	Flügel	Flügel		
	vluchteling	Flüchtling	Flüchtling		
	voorbereiding	Vorbereitung	Vorbereitung		
	vrouw	Frau	Frau		
	water	Wasser	Wasser	water	
	werk	Arbeit	Werk	work	
	zaak	Sache	Sache		
	zaal	Saal	Saal		salle
	lichaam	Körper	(profielwoord)		
	meisje	Mädchen	(profielwoord)		
	voorwerp	Gegenstand	(profielwoord)		
	wet	Gesetz	(profielwoord)		
FR	aai	Ei	Ei	egg	
	behandeling	Behandlung	Behandlung		
	bibleteek	Bibliothek	Bibliothek		bibliothèque
	budzet	Budget	Budget	budget	budget
	doarp	Dorf	Dorf		
	dyk	Deich	Deich	dyke	digue
	eilân	Insel; Eiland	Eiland	island	île
	feest	Fest	Fest	feast	fête
	ferdigening	Verteidigung	Verteidigung		
	ferplichting	Verpflichtung	Verpflichtung		
	fertrouwe	Vertrauen	Vertrauen		
	ferwizing	Verweisung	Verweisung		
	frede	Frieden	Frieden		
	freondin	Freundin	Freundin	friend	
	garaazje	Garage	Garage	garage	garage

Taal	Stimulus	Duitse vertaling	Duits cognaat	Engels cognaat	Frans cognaat
	geslacht	Geschlecht	Geschlecht		
	groep	Gruppe	Gruppe	group	groupe
	haven	Hafen	Hafen		
	klas	Klasse	Klasse	class	classe
	koar	Chor	Chor	choir	chœur
	masine	Maschine	Maschine	machine	machine
	materiaal	Material	Material	material	matériel
	namme	Name	Name	name	nom
	oanfiering	Anführung	Anführung		
	oplossing	Auflösung; Lösung	Auflösung		
	organisaasje	Organisation	Organisation	organisation	organisation
	parlemint	Parlament	Parlament	parliament	parlement
	posysje	Position	Position	position	position
	punt	Punkt	Punkt	point	point
	regearing	Regierung	Regierung		
	rivier	Fluss		river	rivière
	sintimeter	Zentimeter	Zentimeter	centimeter	centimètre
	sintrum	Zentrum	Zentrum	centre	centre
	situaasje	Situation	Situation	situation	situation
	slang	Schlange	Schlange		
	stim	Stimme	Stimme		
	stoarm	Sturm	Sturm	storm	
	súkelade	Schokolade	Schokolade	chocolate	chocolat
	tiisdei	Dienstag	Dienstag	Tuesday	
	ungelok	Unglück	Unglück		
	ontwikkeling	Entwicklung	Entwicklung		
	útspraak	Aussprache	Aussprache		
	woartel	Wurzel	Wurzel		
	wyn	Wein	Wein	wine	vin
	ynfloed	Einfluss	Einfluss	influence	influence
	belied	Politik	(profielwoord)		
	buert	Nachbarschaft	(profielwoord)		
	gat	Loch	(profielwoord)		
	heit	Vater	(profielwoord)		
	lawaai	Lärm	(profielwoord)		
SE	anslutning	Anschluss	Anschluss		
	applåd	Applaus	Applaus	applause	applaudissement
	balans	Balance	Balance	balance	balance
	blad	Blatt	Blatt		
	blod	Blut	Blut	blood	
	bröst	Brust	Brust	breast	

FACTOREN BIJ HET HERKENNEN VAN COGNATEN IN ONBEKENDE TALEN

Taal	Stimulus	Duitse vertaling	Duits cognaat	Engels cognaat	Frans cognaat
	chans	Chance	Chance	chance	chance
	cykel ¹⁴	Fahrrad; Zyklus	Zyklus	cycle	cycle
	eftermiddag	Nachmittag	Nachmittag	afternoon	
	elev	Schüler			élève
	fjäder	Feder	Feder	feather	
	frihet	Freiheit	Freiheit	freedom	
	förbättring	Verbesserung	Verbesserung		
	förhandling	Verhandlung	Verhandlung		
	förmåga	Vermögen	Vermögen		
	gemenskap	Gemeinschaft	Gemeinschaft		
	hår	Haar	Haar	hair	
	intryck	Eindruck	Eindruck		
	invånare	Einwohner	Einwohner		
	konst	Kunst	Kunst		
	krig	Krieg	Krieg		
	makt	Macht	Macht		
	morgon	Morgen	Morgen	morning	
	mur	Mauer	Mauer		mur
	nivå	Niveau	Niveau		niveau
	näsa	Nase	Nase	nose	nez
	omgivning	Umgebung	Umgebung		
	ordning	Ordnung	Ordnung	order	ordre
	press	Presse	Presse	press	presse
	resa	Reise	Reise		
	riktning	Richtung	Richtung		
	skada	Schaden	Schaden		
	stjärna	Stern	Stern	star	star
	strävan	Streben	Streben		
	succé	Erfolg		success	succès
	sång	Lied		song	
	tunga	Zunge	Zunge	tongue	
	tält	Zelt	Zelt		
	undersökning	Untersuchung	Untersuchung		
	universitet	Universität	Universität	university	université
	utbildning	Ausbildung	Ausbildung		
	utland	Ausland	Ausland		
	väg	Weg	Weg	way	
	öga	Auge	Auge	eye	
	ögonblick	Augenblick	Augenblick		
	bonde	Bauer	(profielwoord)		
	hav	Meer	(profielwoord)		
	helg	Wochenende	(profielwoord)		

Taal	Stimulus	Duitse vertaling	Duits cognaat	Engels cognaat	Frans cognaat
	önskan	Wunsch	(profielwoord)		
	skäl	Grund	(profielwoord)		



© 2013 Vanhove & Berthele / Amsterdam University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.