TNTL 125 (2009) 324-344

Renée van Bezooijen, Charlotte Gooskens en
Sebastian Kürschner

Wat weet de Nederlander van de herkomst van Nederlandse woorden?

Abstract – There have been quite a few studies exploring popular ideas and knowledge of variation in the Dutch language area. However, emphasis has always been on the geographic dimension, historic variation remaining underexposed. In the present article we present the results of an investigation which focussed on the knowledge that non-linguists possess of the origin of high-frequent words in present-day spoken Dutch. We hypothesized that loans of English origin would be better recognized than loans from other languages because they are more recent and have been less adapted to the Dutch language. This hypothesis was confirmed. Although words of English origin constitute a relatively small portion of the high-frequent part of the lexicon of present-day spoken Dutch, they are far better recognized than words of German, French and Latin origin.

1 Probleemstelling

De laatste jaren is er steeds meer belangstelling gekomen voor de kennis en ideeën van 'gewone' taalgebruikers over taalvariatie, taalvariëteiten en taalvarianten. Internationaal gezien is het vooral Preston die zich hiermee bezighoudt en die het onderzoeksterrein, dat hij Perceptual Dialectology noemt, vorm heeft gegeven. Centrale onderzoeksvragen binnen de perceptual dialectology zijn:

> What do nonspecialists have to say about variation? Where do they believe it comes from? Where do they believe it exists? What do they believe is its function?' (Preston 1999: xxv).

Deze 'folk characterizations' kunnen vervolgens worden vergeleken met 'scientific characterizations' om te zien wat de overeenkomsten zijn en wat de verschillen.

Volgens Preston liggen de wortels van de perceptual dialectology in het Nederlandstalige gebied. Vooral aan het werk van de Nederlandse dialectoloog Weijnen wordt door hem een belangrijke plaats toegekend in verband met de toepassing van de zogenaamde pijltjesmethode als een alternatieve, niet-wetenschappelijk gefundeerde manier om zicht te krijgen op de indeling van dialectgebieden. Als basis voor de pijltjesmethode dienen de antwoorden van dialectsprekers op de vraag: 'In welke plaats(en) in uw gebied spreekt men hetzelfde of ongeveer hetzelfde dialect als u zelf spreekt?' Vanuit de woonplaatsen van de informanten worden pijlen getekend naar alle plaatsen waar volgens hun mening (ongeveer) hetzelfde dialect wordt gesproken. Zo ontstaat een kaart die de dialect-geografische realiteit representeert zoals die door dialectgebruikers wordt waargenomen. De gebieden waarbinnen plaatsen onderling worden verbonden door pijlen vormen waargenomen groepen van dialecten en de pijlloze gebieden vormen waargenomen dialectgrenzen. Deze perceptuele dialectrealiteit kan vervolgens worden vergeleken met traditionele dialectindelingen op basis van isoglossen afkomstig van professionele

dialectologen. De eerste dialectkaart die volgens de pijltjesmethode tot stand is gekomen, is die van Weijnen uit 1946 van de provincie Noord-Brabant (herdrukt in Weijnen 1966, p. 198). Ook later is in Nederland onderzoek gedaan naar de kennis en ideeën die mensen hebben over de afstanden tussen dialecten en de herkomst van sprekers van dialecten en accenten. Voorbeelden zijn het werk van Van Hout & Münstermann (1988), Gooskens (1997), Van Bezooijen & Ytsma (2000), Van Bezooijen & Heeringa (2006) en Giesbers (2008).

Wat opvalt is dat het Nederlandse perceptuele onderzoek zich tot nu toe uitsluitend heeft gericht op de kennis en ideeën van taalkundige leken over de rol en herkenbaarheid van *geografisch* bepaalde taalkundige kenmerken, dat wil zeggen op de regionaal bepaalde variatie die synchroon in het Nederlands aanwezig is. Er is geen onderzoek gedaan naar de herkenbaarheid van *historisch* bepaalde taalkundige kenmerken, dat wil zeggen naar het bewustzijn bij leken van de diachrone dimensie van de Nederlandse taal. Het doel van het hier gepresenteerde onderzoek is deze leemte te vullen. Wij richten ons op de samenstelling van het hedendaagse lexicon van de Nederlandse standaardtaal en willen door middel van een schriftelijke enquête achterhalen in hoeverre niet academisch-taalkundig geschoolde personen een onderscheid kunnen maken tussen erfwoorden en leenwoorden, en, binnen de groep van leenwoorden, tussen woorden die afkomstig zijn uit het Frans, Latijn, Duits en Engels. We maken daarbij nog een verschil tussen personen met en zonder een gymnasiale opleiding en tussen qua leeftijd onderscheiden personen met en zonder een afgeronde middelbare school opleiding.

Engelse woorden zijn over het algemeen veel recenter opgenomen in de Nederlandse taal dan woorden met een Franse, Latijnse of Duitse herkomst. De meeste Engelse leenwoorden dateren van na 1840 en pas na de tweede wereldoorlog is het lenen van Engelse woorden in een stroomversnelling geraakt. Van der Sijs (1996) noemt als belangrijkste oorzaak de positieve houding tegenover de bevrijders uit Engeland, Canada en de vs, en de hiermee samenhangende modelvorming van de *American way of life.* Ook andere factoren spelen volgens Van der Sijs een rol: het Engels werd steeds meer de voertaal van internationale organisaties en de wetenschap, Engels werd een verplicht schoolvak en Engels werd steeds meer gebruikt in moderne technieken, zoals de computer.

Vergeleken met Engelse leenwoorden zijn Latijnse, Franse en Duitse leenwoorden gemiddeld genomen vele eeuwen ouder. Dit verschil in ouderdom heeft taalkundige consequenties. Nadat een woord geleend is, maakt het vaak dezelfde fonetische, fonologische, orthografische en morfologische veranderingen door als de endogene woorden van de ontlenende taal (Van der Sijs 1996, p. 46). Hoe vroeger het woord in het Nederlands is opgenomen, hoe groter de kans dat zijn vreemde origine niet meer uit de gesproken en geschreven vorm van het woord is af te leiden. De vreemde herkomst kan op den duur onzichtbaar worden. Tegen deze achtergrond verwachten wij dat niet-taalkundig geschoolde Nederlanders makkelijker de herkomst van Engelse leenwoorden zullen herkennen dan de herkomst van Franse, Latijnse en Duitse leenwoorden. In de volgende sectie 2 zullen we de methode beschrijven waarmee we deze hypothese hebben getoetst. In sectie 3 presenteren we de resultaten van het onderzoek. We sluiten af met een korte discussie in sectie 4.

2 Methode

Om de kennis van de samenstelling van het Nederlandse lexicon te onderzoeken hebben we gebruik gemaakt van een schriftelijke enquête die in het voorjaar van 2007 is uitgevoerd.

2.1 Proefpersonen

De enquête is afgenomen bij 103 proefpersonen, 44 mannen en 59 vrouwen. De gemiddelde leeftijd bedroeg 30.6 jaar. De jongste proefpersoon was 15 jaar, de oudste 68 jaar. We hebben de proefpersonen onderverdeeld naar de vordering en aard van hun studie. De jongere groep (15 tot 19 jaar, gemiddeld 17.1) zat nog op de middelbare school (vwo), terwijl de oudere groep (20 tot 68 jaar, gemiddeld 41.7 jaar) bezig was met een studie op het hbo of wo of een studie op dit niveau had afgerond. Hierbinnen hebben we een onderscheid gemaakt tussen proefpersonen met Latijn in hun vakkenpakket en proefpersonen die geen Latijn hadden (gehad). We hadden het idee dat kennis van het Latijn de determinering van de herkomst van woorden zou vergemakkelijken. De precieze aantallen proefpersonen in elk van de vier groepen zijn te vinden in Tabel 1.

Tabel 1 Aantal proefpersonen (totaal en uitgesplitst naar man en vrouw) in elk van de vier groepen

		Opleiding	
		vwo	**post-vwo**
Latijn	Niet	19 (10m en 9v)	37 (15m en 22v)
	Wel	27 (7m en 20v)	20 (12m en 8v)

2.2 Taak

De proefpersonen kregen schriftelijk een lijst van hoog-frequente woorden uit het hedendaags gesproken Nederlands aangeboden waarvan door ons objectief de herkomst was vastgesteld. De proefpersonen dienden aan te geven uit welke taal ze dachten dat de woorden afkomstig waren. De selectie van deze woorden is via een aantal stappen tot stand gekomen. Deze stappen zullen hieronder worden beschreven.

2.3 Spraakmateriaal

2.3.1 Basis

In algemene zin geldt dat de aard van het lexicon sterk wordt bepaald door stijl. In een informele stijl worden andere woorden gebruikt, met een andere herkomst, dan in een formele stijl. Om deze lexicale verscheidenheid goed tot zijn recht te laten komen in de enquête, hebben we woorden geselecteerd uit twee stilistisch onderscheiden spraakverzamelingen.

Enerzijds hebben we gebruik gemaakt van de component 'Spontane conversaties' ('Face-to-face interactions') van het Corpus Gesproken Nederlands (CGN),

aangelegd in de periode 1998-2004.[1] Dit onderdeel bevat gesprekken tussen vrienden en bekenden die door hen zelf zijn opgenomen in de huiselijke kring. Er was geen interviewer aanwezig, en stilistisch kan de spraak gekenmerkt worden als informeel. De component 'Spontane conversaties' bevat in totaal meer dan twee en een half miljoen woorden. De sprekers zijn verdeeld over de twee geslachten, verschillende leeftijdscategorieën en verschillende regio's. Tweederde van het materiaal is afkomstig uit Nederland en een derde uit Vlaanderen.

Anderzijds hebben we gebruik gemaakt van spraakmateriaal afkomstig uit vergaderingen van het Europese Parlement (Europarl Corpus of European Parliament Proceedings) in de eerste maanden van 2000.[2] Het gaat om monologen van sprekers zowel als van de voorzitters. Deze spraak kan als formeel worden gekarakteriseerd. Europarl bevat spraak van elf Europese talen, per taal gemiddeld 28 miljoen woorden. Voor het huidige onderzoek hebben we gebruik gemaakt van iets minder dan een miljoen woorden afkomstig van Nederlandse en Vlaamse sprekers, alsook vertalingen door Nederlandse en Vlaamse tolken van de spraak van anderstalige sprekers.

2.3.2 Woordfrequentie

Ten behoeve van de enquête hebben we uit de twee materiaalverzamelingen een selectie gemaakt. Het eerste selectiecriterium was woordfrequentie. Uit de spontane conversaties van het CGN hebben we de 1500 meest frequente woorden geselecteerd. Het betreft de frequentie van lemma's, wat wil zeggen dat de frequenties van bijvoorbeeld *huis*, *huisje* en *huizen* zijn samengevoegd. Het lemma wordt in onze materiaalverzameling vertegenwoordigd door de enkelvoudige, niet-verkleinde vorm *huis*. Werkwoorden worden vertegenwoordigd door de infinitiefvorm.

Uit de 1500 meest frequente lemma's in het CGN hebben we vervolgens alle persoonsnamen, plaatsnamen en tussenwerpsels (bijvoorbeeld *hoor*, *enfin*, *uhm*, *oh*, *pf*) verwijderd. Daarnaast hebben we transparante samenstellingen, dat wil zeggen samenstellingen waarvan de bijdrage van de delen aan het geheel ook voor een niet taalkundig geschoolde taalgebruiker helder is, opgesplitst en de samenstellende delen afzonderlijk opgenomen. Het kan immers voorkomen dat het ene deel een erfwoord is en het andere een leenwoord. Voorbeelden van opgesplitste samenstellingen zijn *badkamer*, *vrijdagavond*, *schoonmaken* en *lesgeven.* Samenstellingen met een voorzetsel, zoals *afstuderen, doorgaan* en *nadenken,* hebben we niet opgesplitst. Zo bleven er voor de spontane conversaties uit het CGN 1308 woorden over die als informeel kunnen worden gekarakteriseerd.

Dezelfde werkwijze hebben we toegepast op het Europarl-corpus. Hier bleven er, na verwijdering van namen en tussenwerpsels en opsplitsing van samenstellingen, 1400 hoogfrequente woorden over. Deze hebben overwegend een relatief formeel karakter.

Opgeteld over de twee materiaalverzamelingen komt men dan tot 2708 woorden. Er waren echter 556 woorden die in beide corpora voorkwamen (stijlneutraal). Het totale bestand *verschillende* woorden bestond dus uit 2152 woorden. De verdeling naar stijl is voor de overzichtelijkheid nog eens aangegeven in Tabel 2.

1 Zie http://lands.let.kun.nl/cgn/home.htm
2 Zie http://www.statmt.org/europarl/

Tabel 2 Verdeling van de woorden naar stijl

Stijl	Aantal woorden
Informeel (alleen CGN)	752
Formeel (alleen Europarl)	844
Stijlneutraal (zowel in CGN als Europarl)	556
Totaal	2152

2.3.3 Codering

Alle 2152 woorden hebben we met de hand gecodeerd met betrekking tot de volgende informatie:

1 Woordklasse (zelfstandig naamwoord, werkwoord, bijvoeglijk naamwoord, bijwoord, lidwoord, voorzetsel, voornaamwoord, telwoord en voegwoord).
2 Woordtype: inhoudswoord (de eerste vier woordklassen) of functiewoord (de laatste vijf woordklassen).
3 Erfwoord of leenwoord.
4 Indien leenwoord: taal waaraan het rechtstreeks is ontleend.

De informatie met betrekking tot (3) en (4) is in eerste instantie overgenomen uit het *Etymologisch Woordenboek. De herkomst van onze woorden* (Van Veen & Van der Sijs 1997). Als dit geen resultaat opleverde, hebben we onze toevlucht genomen tot het *Leenwoordenboek. De invloed van andere talen op het Nederlands* (Van der Sijs 1996).

2.3.4 Samenstelling

Aan de hand van de codering kon het spraakmateriaal nader worden geanalyseerd. Uit de analyse kwam naar voren dat het spraakcorpus uit 1969 inhoudswoorden en 183 functiewoorden bestond. De onderverdeling naar erfwoord en leenwoord liet het volgende beeld zien: 1541 erfwoorden (71.6%) en 611 leenwoorden (28.4%). In de verdere analyses van de leenwoorden concentreren we ons op de inhoudswoorden, omdat er zich onder de functiewoorden vrijwel geen leenwoorden bevonden. De bijdrage van verschillende talen aan het totale aantal geleende inhoudswoorden is weergegeven in Tabel 3.

Tabel 3 Aandeel van verschillende talen in de geleende inhoudswoorden

Taal	Aantal	Percentage
Frans	351	58.5
Latijn	152	25.3
Duits	43	7.2
Engels	29	4.8
Grieks	4	0.7
Onbekend	8	1.3
Anders	13	2.2
	600	100

In Tabel 3 is te zien dat meer dan de helft van de leenwoorden in ons corpus afkomstig is uit het Frans. Daarnaast zijn ook redelijk wat woorden, ongeveer een kwart, rechstreeks afkomstig uit het Latijn. Deze twee Romaanse talen samen zijn verantwoordelijk voor 83.8% van alle leenwoorden. Op de derde plaats komt het Duits, met een aandeel van 7.2%, en op de vierde plaats komt het Engels, met 4.8%. Extrapolerend betekent dit dat nog niet één op de twintig leenwoorden in het hedendaags gesproken Nederlands van Engelse herkomst is. Het aandeel Engelse leenwoorden op het totaal van alle inhoudswoorden, dus exogeen en endogeen samen, bedraagt in onze materiaalverzameling 1.5%.

2.3.5 Aangeboden woorden

Zoals boven is aangegeven hebben we in de vragenlijst alleen inhoudswoorden opgenomen, dat wil zeggen bijvoeglijke naamwoorden, zelfstandige naamwoorden, bijwoorden en werkwoorden. Qua herkomst van de leenwoorden hebben we ons beperkt tot het Latijn, Frans, Engels en Duits. Woorden met een andere herkomst kwamen in ons materiaal bijna niet voor. Het betreft in alle gevallen de taal van *directe* ontlening. In principe boden we in de vragenlijst tien woorden per cel (combinatie van stijl en taal van directe ontlening) aan. Als er meer dan tien woorden beschikbaar waren in ons materiaal, hebben we er willekeurig tien geselecteerd. Als er minder dan tien beschikbaar waren, hebben we alle beschikbare woorden opgenomen. Alleen voor het Duits en Engels was het aantal van tien niet voor elk van de drie stijlen haalbaar. Voor het Duits zijn er in totaal 29 woorden aangeboden en voor het Engels 18, zoals te zien is in Tabel 4.

Tabel 4 Aantal woorden voor elke combinatie van stijl en taal van herkomst

Stijl	Erfwoord	Leenwoord				Totaal
	Nederlands	Duits	Engels	Frans	Latijn	
Informeel	10	9	10	10	10	49
Formeel	10	10	5	10	10	45
Neutraal	10	10	3	10	10	43
Totaal	30	29	18	30	30	137

In de enquête zijn in totaal 137 woorden afgevraagd (zie Bijlage). Deze zijn door elkaar in willekeurige volgorde aangeboden op papier. Bij elk woord is gevraagd: Uit welke taal denk je dat dit woord afkomstig is, Nederlands, Latijn, Frans, Engels, Duits of een andere taal? De proefpersonen hadden ook de mogelijkheid 'ik heb geen idee' aan te kruisen. Het invullen van de enquête duurde gemiddeld zo'n twintig minuten.

3 Resultaten

3.1 Alle proefpersonen samen

Allereerst is het interessant om vast te stellen hoe vaak de proefpersonen correct het globale onderscheid hebben gemaakt tussen erfwoord en leenwoord. De Nederlandse erfwoorden zijn in 69% van de gevallen correct als inheemse woorden herkend. De verschillende typen leenwoord zijn gemiddeld in 73% van de gevallen als niet-inheemse woorden herkend. De niet-Nederlanse herkomst van de Duitse leenwoorden is blijkbaar het minst transparant (43%), die niet-Nederlandse herkomst van de Engelse (91%), Franse (80%) en Latijnse (77%) leenwoorden wordt relatief goed door de proefpersonen herkend. Dat de Duitse leenwoorden moeilijk te onderscheiden zouden zijn van de Nederlandse erfwoorden en de Engelse juist makkelijk vloeit voort uit de in de inleiding verwoorde hypothese, gebaseerd op de mate van aanpassing aan het Nederlands in samenhang met het tijdstip van ontlening. Dat, naast het Engels, ook de vreemde herkomst van de Franse en Latijnse leenwoorden zo goed wordt herkend, is echter wat onverwacht.

Om enig inzicht te krijgen in de oorzaak van de relatief goede herkenbaarheid van de uitheemse oorsprong van de Franse en Latijnse leenwoorden hebben we hun vormelijke kenmerken vergeleken met die van de Nederlandse erfwoorden (zie Bijlage). Hierbij vielen ons drie dingen op.

Ten eerste is er een aanmerkelijk verschil in de frequentie van voorkomen van de letter *c*. In geen van de aangeboden Nederlandse erfwoorden komt de letter *c* voor, terwijl negen van de dertig Franse woorden (30%) en tien van de dertig Latijnse woorden (33%) een *c* bevatten. De letter *c* wordt door Van Heuven, Neijt & Hijzelendoorn (1995) als een 'exotische grafie' omschreven.

Ten tweede hebben de Franse en Latijnse leenwoorden nogal wat affixen die de proefpersonen mogelijkerwijs, generaliserend vanuit hun vreemde-talenkennis, als niet-Nederlands hebben herkend, zoals de prefixen *inter-*, *para-*, *ab-*, *con-* en *pro-*, en de suffixen *-aal*, *-iteit* en *-ie*.

Ten derde is er een opvallend verschil in de aard van de klinkers in de meerlettergrepige woorden. Van de 26 Franse meerlettergrepige woorden bevatten er 23 meer dan een volle klinker (88%) en van de 24 Latijnse meerlettergrepige woorden 21 (88%). Van de twintig Nederlandse meerlettergrepige woorden hebben er slechts drie meer dan een volle klinker (15%), en in alledrie de gevallen is dit te wijten aan volle klinkers in de duidelijk te herkennen affixen *aan-* (*aangenaam*) en *-heid* (*gelegenheid*, *overheid*). Volgens Van Heuven, Neijt en Hijzelendoorn (1995) bevat een (van affixen ontdaan) Nederlands erfwoord niet meer dan een lettergreep met een volle klinker of tweeklank.

Wij hebben het idee dat de proefpersonen op een of andere manier kennis hebben van de etymologische betekenis van de drie bovengenoemde woordkenmerken en er gebruik van hebben gemaakt bij hun indeling in erfwoord en leenwoord. Wij denken, met andere woorden, dat de letter *c*, affixen als *inter-* en *-aal*, en meerdere volle klinkers als aanwijzingen zijn gezien voor een uitheemse herkomst. Als proef op de som hebben we twee dingen gedaan.

Aan de ene kant hebben we gekeken naar de Franse en Latijnse woorden die geen van de drie beschreven kenmerken bezitten. Onze voorspelling is dat dit het

voor de proefpersonen moeilijk maakt om van deze woorden de uitheemse herkomst te bepalen en dat de percentages correct uitheems dus relatief laag zullen zijn. Bij 'uitheems' hebben we (in tegenstelling tot de percentages in de eerste alinea van deze sectie) het Duits niet meegerekend, omdat Duits de genoemde, Germaanse kenmerken gemeen heeft met het Nederlands. Bij 'uitheems' hebben we naast Romaanse (Franse of Latijnse) herkomst wel het Engels meegerekend, omdat veel woorden van Romaanse oorsprong via het Engels het Nederlands hebben bereikt. Concreet blijken er zes Romaanse meerlettergrepige woorden te zijn aangeboden die geen van de drie genoemde kenmerken hebben. Het gaat om de Franse woorden *letter* (73% correct uitheems), *simpel* (52%) en *fraude* (81%), en de Latijnse woorden *koken* (23%), *vieren* (3%) en *tafel* (34%). Voor de Latijnse woorden lijken onze ideeën te worden bevestigd, hier zijn de percentages uitheems inderdaad relatief laag. Voor de Franse woorden lijkt er echter iets anders aan de hand te zijn, hier zijn de percentages, ondanks de afwezigheid van de drie kenmerken, toch relatief hoog. Dit kan te maken hebben met het feit dat de Franse woorden in vrijwel dezelfde vorm in het hedendaags Engels/Frans voorkomen en de proefpersonen op die manier bekend zijn. Nader onderzoek is hier zeker op zijn plaats.

Als tweede test hebben we juist gekeken naar de woorden uit het Frans en Latijn die alledrie de door ons beschreven kenmerken bezitten. Naar onze verwachting zouden deze juist goed als uitheems herkend moeten zijn en zouden de percentages uitheemse herkomst hier dus relatief hoog moeten zijn. Het gaat om de Franse woorden *accepteren* (90% correct uitheems), *constateren* (83%), *document* (93%) en *activiteit* (90%), en de Latijnse woorden *conferentie* (98%) en *commissaris* (84%). Onze verwachting wordt bevestigd. Franse en Latijnse woorden die gekenmerkt worden door de letter c, een Romaans affix en meerdere volle klinkers worden door veel proefpersonen correct als uitheems geïnterpreteerd. Ook hier geldt dat de precieze rol van de drie kenmerken, de rol van eventuele andere kenmerken en de rol van algemene talenkennis verder onderzocht zouden moeten worden.

In dit artikel gaat onze aandacht met name uit naar het percentage correcte herkenning van de specifieke taal van ontlening, en daartoe zullen we ons verder ook beperken. Het totale percentage woorden waarvan de speficieke taal van directe ontlening correct door de proefpersonen is aangegeven bedraagt 50%, dat is dus de helft. In Tabel 5 zijn de resultaten gespecificeerd per taal te vinden, met de percentages correcte identificatie op de diagonaal en daarbuiten de verwarringen met andere talen. Er was ook nog een responsiecategorie 'andere taal', maar deze hebben we weggelaten omdat er weinig gebruik van is gemaakt. Dit verklaart ook waarom de percentages in Tabel 5 en 6 niet altijd optellen tot 100%.

Uit Tabel 5 kan worden opgemaakt dat de herkenning van de Nederlandse erfwoorden redelijk goed gaat (69%), maar dat er wel sprake is van een sterke response bias. Als de responsies van de proefpersonen gelijkmatig over alle talen verdeeld waren geweest, zou dit uitkomen op 20% per taal. In de onderste rij van Tabel 5 is echter te zien dat de proefpersonen bovengemiddeld vaak (38%) voor het Nederlands als taal van herkomst hebben gekozen en ondergemiddeld weinig (tussen de 13% en 18%) voor de andere vier talen. Mensen overschatten dus het aandeel erfwoorden in het hedendaags gesproken Nederlands en onderschatten

het aandeel woorden dat aan andere talen is ontleend. Als een bepaalde taal, in dit geval het Nederlands, bovengemiddeld vaak wordt gekozen, is de kans dat de responsie correct is natuurlijk ook groter.

Wat betreft de leenwoorden blijkt de herkomst van het Engels veel vaker (64%) correct te worden geïdentificeerd dan die van de andere leentalen, die alle drie rond de 30% zitten: Duits 29%, Frans 32% en Latijn 34%. Met andere woorden, bijna twee derde van de Engelse leenwoorden worden als zodanig door de proefpersonen herkend, terwijl slechts één derde van de leenwoorden uit de andere drie talen als zodanig herkenbaar zijn.

Tabel 5 Responsies per taal en opgeteld over talen. Op de diagonaal (in donkergrijs) de percentages correct geïdentificeerde woorden. In lichtgrijs: frequente verwarringen ≥20%.

		Responsie				
		Ned	Duits	Engels	Frans	Latijn
Stimulus	Ned	**69**	20	6	1	2
	Duits	57	**29**	4	3	5
	Engels	9	3	**64**	10	13
	Frans	20	5	18	**32**	25
	Latijn	23	8	17	17	**34**
	Totaal	38	14	18	13	16

Als de herkomst van een woord niet correct is geïdentificeerd, met welke herkomst is er dan verwarring opgetreden? In Tabel 5 zijn in de buitendiagonale cellen de percentages verwarring te vinden. In onze bespreking van de resultaten houden we enigszins willekeurig een drempel aan van 20%. Verwarringen die 20% of vaker voorkomen, noemen we frequent. Deze zijn in Tabel 5 aangegeven in lichtgrijs. Verreweg de vaakst voorkomende verwarring wordt gevormd door de Duitse leenwoorden die voor Nederlandse erfwoorden worden aangezien (57%). Blijkbaar zijn de Duitse leenwoorden in veel gevallen niet te onderscheiden van Nederlandse woorden. Andersom worden ook veel Nederlandse erfwoorden ten onrechte als Duits gezien (20%). Naast Duitse leenwoorden worden ook Franse (20%) en Latijnse (23%) leenwoorden geregeld geïnterpreteerd als Nederlandse erfwoorden. Ten slotte worden Franse leenwoorden vaak (25%) geïdentificeerd als Latijnse leenwoorden. De relatief weinig frequente incorrecte responsies voor de Engelse leenwoorden zijn niet geconcentreerd in een enkele taal maar gespreid.

Omdat het Engels zo opvalt door het hoge pecentage correcte identificaties, leek het ons de moeite waard de responsies voor het Engels gedetailleerder te bekijken. In Tabel 6 hebben we de percentages afzonderlijk voor ieder Engels woord aangegeven, geordend van het hoogste naar het laagste percentage correcte herkenning.

In Tabel 6 is te zien dat de herkenbaarheid van de verschillende Engelse leenwoorden sterk uiteenloopt. Eén woord is door iedereen als Engels leenwoord herkend, namelijk *mail*. Nog zes andere woorden zijn ook door een grote meerderheid (> 80% van de proefpersonen) correct als Engels geïdentificeerd, namelijk *shit*, *internet*, *team*, *club*, *tanker* en *weekend*.

Tabel 6 Percentages correcte herkenning en verwarringen afzonderlijk voor alle Engelse stimuluswoorden. De woorden zijn geordend van best herkenbaar naar slechtst herkenbaar als Engels leenwoord. De taal die het vaakst gerespondeerd is per woord is vet afgedrukt.

	Responsie					
Woord	Engels	Nederlands	Duits	Frans	Latijn	Anders
Mail	**100**	0	0	0	0	0
shit	**98**	0	0	1	1	0
internet	**94**	0	1	0	4	1
team	**92**	3	0	2	3	0
club	**88**	0	1	9	2	0
tanker	**84**	6	7	3	1	0
weekend	**83**	6	4	3	4	0
partner	**73**	7	7	9	3	2
sport	**70**	12	3	8	7	0
plannen	**63**	28	4	3	3	0
starten	**55**	36	3	0	3	3
film	**53**	12	6	21	8	0
bus	**47**	22	6	17	6	2
video	42	1	0	5	**52**	0
internationaal	**34**	8	3	24	30	2
infrastructuur	28	13	2	7	**48**	3
foto	23	4	2	**48**	21	3
structureel	18	11	7	24	**37**	3

De onderste vijf woorden in Tabel 6, dat wil zeggen *video*, *internationaal*, *infrastructuur*, *foto* en *structureel*, vormen een aparte categorie. Al deze woorden zijn om begrijpelijke redenen, en in zekere zin ook correct, vaak als Frans of Latijn (en af en toe ook als Grieks) genoteerd. Wellicht ten overvloede willen wij er hier nog eens op wijzen dat wij een zeer strenge norm hebben gehanteerd en een responsie alleen goed hebben gerekend als de proefpersoon correct de taal van *directe* lening aangaf. Veel Engelse woorden zijn echter op hun beurt aan het Frans ontleend en gaan via het Frans weer terug op het Latijn. Het is niet zo vreemd dat veel proefpersonen niet op de hoogte zijn van de exacte route die woorden hebben afgelegd en een taal hebben genoemd waaraan het woord *indirect* is ontleend. Helemaal fout kun je een dergelijke keuze niet vinden. Ook taalwetenschappers hebben hier vaak moeite mee (Van der Sijs 1996, p. 46). Een bijzonder geval in deze context is het woord *foto*. In Van Veen en Van der Sijs (1997), dat wij primair als basis hebben gebruikt voor de bepaling van de etymologie van onze woorden, wordt bij *foto* aangegeven dat het om een Engels leenwoord gaat, dat is afgeleid van het woord *photo*, verkort uit *photography*. Bij *fotografie* staat vervolgens dat het aan het Frans is ontleend en dat het is afgeleid van *photographie*, dat op zijn beurt teruggaat op het Engelse *photography*. Geen wonder dat veel proefpersonen in de enquête bij *foto* hebben gekozen voor Frans eerder dan voor Engels. Deze ingewikkelde situatie geldt vooral voor de laatste vijf woorden in Tabel 6 en heeft veel minder een rol gespeeld bij de overige woorden.

Is er nu een patroon te ontdekken in de responsies van de proefpersonen? Is het duidelijk waarom sommige woorden zo veel vaker correct als Engels leenwoord

zijn geïdentificeerd dan andere? Op wat voor informatie zouden de proefpersonen hun oordelen hebben gebaseerd? Wij denken dat de aanwezigheid van de volgende kenmerken van de stimuluswoorden de proefpersonen zullen hebben gestimuleerd om te kiezen voor de responsie 'Engels leenwoord':

1 Gelijkende of identieke woordvorm in het hedendaags geschreven Engels (cognaat).
2 Engels spellingskenmerk, dat wil zeggen een letter of lettercombinatie die wel in het Engels maar (vrijwel) niet in het Nederlands voorkomt, bijvoorbeeld *ai*, *sh* en *ea*.
3 Engelse relatie tussen spelling en uitspraak, dat wil zeggen een letter(combinatie)/klank-relatie die wel in het Engels maar (vrijwel) niet in het Nederlands voorkomt, bijvoorbeeld *ai* uitgesproken als [e:], *sh* als [ʃ], *ea* als [ie:], *a* als [ɛ], *ee* als [i] of *a* uitgesproken als [ɑ].
4 Met de vs of Groot-Brittannië geassocieerd maatschappelijk domein. Ons is geen studie bekend waarin deze associaties zijn onderzocht. We hebben ons daarom gebaseerd op de frequente ontleningsdomeinen die door Van der Sijs (1996) worden genoemd, namelijk 'computer' (p. 322), 'sport' (p. 318) en 'audiovisuele middelen' (p. 319).

In Tabel 7 zijn voor alle Engelse leenwoorden de vier genoemde kenmerken geïnventariseerd. Uit de vierde kolom kan worden opgemaakt dat alle achttien Engelse leenwoorden een cognate vorm hebben in het hedendaags Engels. Het betreft de woordvormen die in de derde kolom gegeven zijn. In de meeste gevallen (twaalf van de achttien woorden) is de Nederlandse geschreven vorm identiek aan de Engelse geschreven vorm (score 1), in de andere gevallen is er sprake van een verschil in schrijfwijze (score 0.5). Bij *foto* is het louter een aanpassing van de spelling (*ph>f*), bij *plannen* en *starten* zijn de woorden aangepast aan de Nederlandse morfologie van de werkwoordsvervoeging (infinitiefvorm op *-en*) en bij *internationaal*, *infrastructuur* en *structureel* zijn op zijn Nederlands gespelde derivationele suffixen aangehecht. Hierbij dient wel te worden aangetekend dat een aantal woorden (om historische redenen) ook in (vrijwel) identieke vorm in het hedendaags Frans voorkomen. Echter, omdat veel Nederlanders weinig of geen Frans kennen, hebben ze hier geen weet van.

De scores voor de afzonderlijke kenmerken alsmede de totaalscore hebben we gecorreleerd met de percentages correcte herkenning. De uitkomsten zijn te vinden in Tabel 8. Het is te zien dat drie van de vier afzonderlijke kenmerken significant correleren met de mate van herkenning. De hoogste correlatie is die met het kenmerk 'cognaat' (r=.72). Dit wil zeggen dat de woorden waarvan de spelling in het Nederlands identiek is aan die in het Engels makkelijker herkend worden dan de woorden waarvan de spelling op een of meer punten is vernederlandst. Eén kenmerk vertoont geen significante correlatie, namelijk 'associatie met Amerikaans / Brits maatschappelijk domein'. Het kan zijn dat dit kenmerk geen rol speelt of dat het kenmerk niet correct is gescoord. Zoals eerder aangegeven is er voor zover wij weten geen empirisch onderzoek gedaan naar deze associaties.

Als laatste analyse hebben we een multipele regressie uitgevoerd. Hiermee kan worden vastgesteld hoeveel van de variantie in de herkenningsscores kan worden verklaard met de vier kenmerken samen. We hebben gekozen voor de stepwise methode.

Tabel 7 Kenmerken van de Engelse leenwoorden. De woorden zijn geordend van best herkenbaar naar slechtst herkenbaar. In de tweede kolom het jaartal / de periode van het eerste geregistreerde voorkomen in het Nederlands zoals genoemd in Van Veen & Van der Sijs (1997); *internet*, *plannen* en *infrastructuur* ontbreken hierin als lemmata.[3] In de derde kolom de corresponderende Engelse vorm. De volgende vier kolommen verwijzen naar de kenmerken 1 tot en met 4 zoals die in de tekst zijn beschreven. Bij 'Engels cognaat' betekent '0.5' dat de geschreven Engelse vorm een of meer verschillen vertoont met de geschreven Nederlandse vorm; '1' geeft aan dat de geschreven vormen in het Engels en Nederlands identiek zijn. In de laatste kolom is voor de overzichtelijkheid het percentage correcte herkenning nog eens overgenomen uit Tabel 6.

Woord	Jaar	Engelse vorm	Eng. cognaat	Eng. Spelling	Eng. spelling-uitspraak correspondentie	Amerikaans /Brits maatsch. domein.	Totaalscore	% correct
mail	1847[4]	mail	1	1	1	1	4	100
shit	na 1950	shit	1	1	1	–	3	98
internet	?	internet	1	–	–	1	2	94
team	901/1925	team	1	1	1	1	4	92
club	1800	club	1	–	–	1	2	88
tanker	1926/1950	tanker	1	–	1	–	2	84
weekend	1901/1925	weekend	1	–	1	–	2	83
partner	1847	partner	1	–	1	–	2	73
sport	1847	sport	1	–	–	1	2	70
plannen (ww)	?	plan	0.5	–	1	–	1.5	63
starten	1919	start	0.5	–	–	–	0.5	55
film	1901/1925	film	1	–	–	1	2	53
bus	1887	bus	1	–	–	–	1	47
video	na 1950	video	1	–	–	1	2	42
internationaal	1847	international	0.5	–	–	–	0.5	34
infrastructuur	?	infrastructure	0.5	–	–	–	0.5	28
foto	1901/1925	photo	0.5	–	–	1	1.5	23
structureel	1926/1950	structural	0.5	–	–	–	0.5	18

Hierbij wordt eerst de hoogstcorrelerende variabele gekozen en daarna worden variabelen toegevoegd als ze een significante bijdrage leveren aan de totale verklaarde variantie. In ons geval werd met de variabelen 'cognaat' (als eerste ingevoerd) en 'Engelse spelling-uitspraak correspondentie' (als tweede ingevoerd) een multipele R verkregen van .84. Dit betekent dat met deze twee kenmerken 66% (R^2) van de variantie in de herkenningssscores kan worden verklaard. Dit is een goed resultaat.

3 In het *Etymologisch Woordenboek* van Van Dale (Van Veen en Van der Sijs 1997: xxviii) worden verschillende redenen genoemd waarom in bepaalde gevallen niet een jaartal van de oudste vindplaats wordt genoemd, maar een periode. Soms kan bijvoorbeeld de tekst waarop de datering is gebaseerd niet exact gedateerd worden. Dit geldt vooral voor Middeleeuwse woorden. De dateringen uit de twintigste eeuw zijn over het algemeen in periodes gegeven omdat van de overvloedige bronnen slechts een klein deel is geraadpleegd. Een exact jaartal zou dan te veel zekerheid suggereren.

4 Deze datering betreft de betekenis 'brievenpost'. De datering in de betekenis 'electronische post' is niet aangegeven, maar deze is natuurlijk veel recenter.

Tabel 8 Correlaties van vier kenmerken van Engelse leenwoorden alsmede de totaalscore met percentages correcte herkenning. * p <.05, ** p < .01, *** p < .001.

Kenmerk	Coëfficiënt	Significantie
Cognaat	.72	***
Spelling	.56	*
Spelling-uitspraak correspondentie	.64	**
Maatschappelijk domein	.22	n.s.
Totaalscore	.78	***

Tabel 9 Percentages correcte herkenning en verwarringen afzonderlijk voor alle Duitse stimuluswoorden (voor een toelichting, zie tabel 6)

	Responsie					
Woord	Duits	Nederlands	Engels	Frans	Latijn	Anders
Sowieso	**71**	26	0	1	1	1
richting	48	**50**	0	0	2	0
erts	**44**	38	6	2	8	2
ongeveer	42	**54**	1	1	2	0
omgeving	41	**54**	1	0	2	2
regelmatig	39	**54**	1	2	4	0
grens	39	**56**	0	0	3	2
ogenblik	38	**59**	0	1	1	1
heersen	37	**55**	0	2	6	0
herinneren	36	**59**	2	0	3	0
praktisch	**36**	10	19	14	18	4
gevaar	35	**63**	0	0	2	0
eenzaam	32	**63**	1	0	2	2
bewust	31	**65**	0	0	2	2
toestand	28	**69**	1	1	2	0
opname	27	**67**	1	1	2	2
straf	24	**63**	2	2	5	4
onderhavig	22	**67**	2	0	3	7
maatregel	20	**72**	2	1	3	2
bevoegd	20	**78**	0	0	1	2
verzamelen	19	**75**	1	0	4	1
wedstrijd	19	**76**	2	1	2	0
invloed	18	**50**	17	10	4	1
bewerkstelligen	18	**76**	2	0	3	1
overigens	18	**75**	1	0	6	0
reageren	16	19	13	16	**31**	4
technisch	15	11	**38**	14	16	1
kast	9	**79**	2	3	7	6
trui	7	**75**	1	11	1	0

Om een meer algemeen beeld te krijgen van met name de taalkundige aspecten die een rol spelen bij de herkenning van leenwoorden, hebben we een vergelijkbare analyse die we bij de Engelse leenwoorden hebben uitgevoerd ook op de Duitse leenwoorden toegepast (Tabel 9 en 10). De Duitse leenwoorden zijn in tegenstelling tot de Engelse als zodanig slecht door de proefpersonen herkend. Het gemiddelde herkenningspercentage voor Duits is 29% tegenover 64% voor Engels. Van de achttien Engelse leenwoorden in Tabel 6 werden er elf door meer dan de helft van de proefpersonen als zodanig herkend, van de 29 Duitse leenwoorden in tabel 9 wordt er slechts één door meer dan de helft van de proefpersonen correct geïdentificeerd, namelijk *sowieso* (71%). Dit is tevens de meest recente Duitse ontlening. In tabel 9 kan worden gezien dat verreweg de meeste Duitse leenwoorden vaker als oorspronkelijk Nederlandse woorden zijn gezien dan als aan het Duits ontleende woorden. Dit was bij geen enkel Engels leenwoord het geval. Een vergelijking van tabel 7 en tabel 10, waarin een aantal kenmerken van de Engelse respectievelijk Duitse leenwoorden is geïnventariseerd, geeft meerdere aanwijzingen waaruit dit verschil kan worden verklaard.

In tabel 10 zijn de resultaten van de analyse van de Duitse leenwoorden te vinden. Deze tabel heeft dezelfde opzet als Tabel 7 voor de Engelse leenwoorden, maar is op twee punten uitgebreid. Ten eerste is er een kolom toegevoegd die aangeeft uit welk soort Duits het leenwoord afkomstig is: Hoogduits of Nederduits. Het Hoogduits werd vanaf de achtste eeuw gesproken ten zuiden van de lijn Keulen-Berlijn, het Nederduits werd gesproken in het gebied van de benedenloop van de grote rivieren Rijn, Elbe, Weser en Ems. Het Nederduits lijkt meer op het Nederlands dan het Hoogduits, omdat beide de Hoogduitse klankverschuiving niet hebben doorgemaakt. Ten tweede is er een kolom toegevoegd die aangeeft of het een leenvorming betreft, dat wil zeggen 'een samenstelling of afleiding die naar het voorbeeld van een andere taal is gevormd' (Van der Sijs 1996, p. 234). Hierbij wordt gebruik gemaakt van corresponderende cognaten.Een voorbeeld is het Duitse *Haushaltung*, dat is omgezet naar het Nederlandse *huishouding*. Voor alle duidelijkheid: leenvormingen dienen te worden onderscheiden van leenvertalingen, waarbij een Duitse vorm is vervangen door een Nederlandse vorm met dezelfde betekenis. Een voorbeeld van een Duitse leenvertaling is *leedvermaak*, dat een vertaling is van *Schadenfreude*. Leenvormingen kwamen bij de Engelse leenwoorden niet voor. Een leenvorming (evenals een leenvertaling) kan de identificatie van de exogene herkomst van een woord bemoeilijken omdat het resulterende woord geheel of gedeeltelijk uit endogene elementen bestaat. Als het Duitse woord *einsam* alleen qua klanken zou zijn aangepast aan het Nederlands, zou het in het Nederlands zoiets als *einzaam* zijn geworden. Het element *ein* is echter vervangen door het Nederlandse element *een*. Het woord krijgt hierdoor een authentiek Nederlands karakter. In Tabel 9 hebben we de woorden die als leenvormingen kunnen worden aangemerkt een negatieve score (-1.00) gegeven, omdat er sprake is van bemoeilijking van de herkenning eerder dan van vergemakkelijking.

Uit Tabel 10 kan worden opgemaakt dat weliswaar vrijwel alle woordvormen van de Duitse leenwoorden in het hedendaags Duits aanwezig zijn (de uitzonderingen zijn *trui* en *kast*, waarvan geen cognate vormen meer bestaan), maar dat slechts vier van de 29 Duitse leenwoorden (13.8%) in identieke vorm in het hedendaags geschreven Duits voorkomen. In het Engels gold dit voor twaalf van

Tabel 10 Kenmerken van de Duitse leenwoorden. De tabel is op een vergelijkbare wijze ingericht als Tabel 7. Er zijn twee kolommen toegevoegd, namelijk een specificatie van het soort Duits (Hoogduits of Nederduits) waartoe het woord oorspronkelijk behoorde en of het een leenvorming betreft. Een leenvorming is als een belemmerende factor beschouwd voor de identificatie van de Duitse herkomst en krijgt daarom een negatieve score.

Woord	Jaar	Duitse vorm	Type Duits	Duitse cognaat	Duitse spelling	Duitse spelling-uitspraak corresp.	Leenvorm	Duits maatschap. domein	Totaalscore	% correct
sowieso	1950	sowieso	H	1	–	1	–	–	2	71
richting	1786	Richtung	H	0.5	–	–	-1	–	-0.5	48
erts	1556	Erz	H	0.5	–	–	–	–	0.5	44
ongeveer	1599	ungefähr	H	0.5	–	–	-1	–	-0.5	42
omgeving	1825	Umgebung	H	0.5	–	–	-1	–	-0.5	41
regelmatig	1661	regelmäßig	H	0.5	–	–	-1	–	-0.5	39
grens	1573	Grenze/ Grenize	H/N	0.5	–	–	–	–	0.5	39
ogenblik	1517	Augenblick	H	0.5	–	–	-1	–	-0.5	38
heersen	1348	herrschen	H	0.5	–	–	–	–	0.5	37
herinneren	1644	erinnern	H	0.5	–	–	-1	–	-0.5	36
praktisch	1840	praktisch	H	1	–	–	–	–	1	36
gevaar	1574	Gefahr	H	0.5	–	–	–	–	0.5	35
eenzaam	1477	einsam	H	0.5	–	–	-1	–	-0.5	32
bewust	1638	bewusst	H	0.5	–	–	–	–	0.5	31
toestand	1648	Zustand	H	0.5	–	–	-1	–	-0.5	28
opname	1828	Aufnahme	H	0.5	–	–	-1	–	-0.5	27
straf	1557	Strafe	H	0.5	–	–	–	–	0.5	24
onderhavig	1818	unterhaben	H	0.5	–	–	-1	–	-0.5	22
maatregel	1734	Maßregel	H	0.5	–	–	-1	–	-0.5	20
bevoegd	1698	befugt	H	0.5	–	–	–	–	0.5	20
verzamelen	1444	versammeln	H	0.5	–	–	–	–	0.5	19
wedstrijd	1650	Wettstreit	H	0.5	–	–	–	–	0.5	19
invloed	1282	Einfluss	H	0.5	–	–	-1	–	-0.5	18
bewerkstelligen	1769	bewerkstelligen	H	1	–	–	–	–	1	18
overigens	1735	übrigens	H	0.5	–	–	-1	–	-0.5	18
reageren	1847	reagieren	H	0.5	–	–	-1	–	-0.5	16
technisch	1778	technisch	H	1	–	–	–	–	1	15
kast	1364	Kasten	H	–	–	–	–	–	0	9
trui	1477/ 1877[5]	Troie	N	–	–	–	–	–	0	7

5 In Van Veen en Van der Sijs (1997) wordt 1477 genoemd bij de vorm *troye* en 1877 bij de vorm *trui*.

de achttien woorden (66.7%). Dit betekent dat de proefpersonen waarschijnlijk minder gauw zullen denken aan het Duits als brontaal. Verder valt op dat geen enkel Duits leenwoord een typisch Duits spellingkenmerk heeft, terwijl dit bij drie (16.7%) van de Engelse woorden wel het geval was. Sterker nog is het verschil met het Engels bij de spelling-uitspraak-relatie. Slechts één van de 29 Duitse leenwoorden (3.4%) heeft een typisch Duitse spelling-uitspraak-correspondentie, namelijk *sowieso*, waarbij de twee *s*'en niet als [s] worden uitgesproken maar als [z]. Van de achttien Engelse leenwoorden hadden er zeven (38.9%) een typisch Engelse spelling-uitspraak-relatie. Vergelijking tussen Tabel 10 en Tabel 7 laat ten slotte zien dat er onder de Duitse leenwoorden veel leenvormingen zijn (14 van de 29 woorden, dat wil zeggen 48.3%), terwijl dit bij de Engelse leenwoorden in het geheel niet voorkwam. Leenvormingen kunnen worden gezien als een aanpassing aan de ontlenende taal die verder gaat dan een aanpassing in de uitspraak. Dit maakt de identificatie van de herkomst moeilijker. Geen van de Duitse leenwoorden riepen bij ons een sterke associatie op met de Duitse cultuur, de meeste Duitse leenwoorden zijn vrij neutraal van inhoud.

Net zoals wij voor de Engelse leenwoorden hebben gedaan, hebben we de kenmerken van de Duitse leenwoorden gecorreleerd met de herkenningsscores. Het betreft in dit geval drie variërende kenmerken, namelijk 'cognaat', 'spelling-uitspraak-correspondentie' en 'leenvorming' en daarnaast de totaalscore. De uitkomsten van de analyse worden gegeven in Tabel 11. Het valt op dat de totaalscore niet significant correleert met het herkenningspercentage, terwijl dit wel geldt voor twee afzonderlijke kenmerken. Het kenmerk 'cognaat' correleert slechts zwak, terwijl 'spelling-uitspraak-correspondentie' een wat hogere correlatie oplevert. We hebben hier typisch met een geval te maken dat één overeenkomend paar gegevens verantwoordelijk is voor een significante correlatie: het woord *sowieso* heeft verreweg het hoogste percentage correcte herkenningen terwijl het het enige woord is dat een niet-Nederlandse, uit het Duits afkomstige spelling-uitspraak-correspondentie heeft (de twee *s*'en worden niet als s uitgesproken maar als z). Aan deze correlatie kan niet veel gewicht worden toegekend. De multipele regressie leverde een R op van .59. De enige opgenomen variabele is 'cognaat'.

Alles in aanmerking nemend verwondert het ons niet dat zo weinig van de Duitse leenwoorden als zodanig door de proefpersonen zijn herkend. Verreweg de meeste Duitse leenwoorden (26 van de 29, d.w.z. 89.6%) zijn (veel) vaker aangezien voor een Nederlands erfwoord dan voor een Duits leenwoord. De enige

Tabel 11 Correlaties van drie kenmerken van Duitse leenwoorden alsmede de totaalscore met percentages correcte herkenning. * p <.05, ** p < .01, *** p < .001.

Kenmerk	Coëfficiënt	Significantie
Cognaat	.38	*
Spelling-uitspraak-correspondentie	.59	***
Leenvorming	-.08	n.s.
Totaalscore	.23	n.s.

uitzonderingen zijn het eerder genoemde *sowieso* en daarnaast *erts*, *praktisch* en *technisch*. De laatste twee medeklinkers van *erts* kunnen als een typisch (maar niet uniek) Duitse consonantcluster worden beschouwd en het suffix *-isch* wordt wanneer het als [iʃ] wordt uitgesproken, soms als sjibolet voor een Nederlandssprekende Duitser gebruikt. Deze Duitse kenmerken kunnen hebben verhinderd dat deze woorden als Nederlandse erfwoorden zijn gezien. Alle andere Duitse leenwoorden hebben een klank- en spellingvorm die niet te onderscheiden is van 'echte' Nederlandse woorden. Het feit dat er vaak sprake is van leenvorming heeft hier sterk aan bijgedragen. Hieraan dient te worden toegevoegd dat de kennis van het Duits bij Nederlanders aanzienlijk geringer is dan die van het Engels. Dit aspect zal zeker ook hebben bijgedragen tot de lagere identificatiescores voor het Duits.

3.2 Effect van de vordering van de studie

In Tabel 12 zijn de responsies uitgesplitst naar opleiding: vwo en post-vwo. Deze opdeling correspondeert met een verschil in leeftijd. De vwo-groep is gemiddeld zo'n 25 jaar jonger dan de post-vwo-groep. Het totale percentage correct voor de vwo proefpersonen is 40.3% en voor de post-vwo-proefpersonen 46.6%. Dit kan erop wijzen dat proefpersonen na hun middelbare schoolopleiding nog bijleren over de herkomst van Nederlandse woorden. Het kan ook zijn dat de kennis van vreemde talen bij mensen met een opleiding vroeger groter was dan nu. Over de hele linie is er een grote gelijkenis in het herkenningspatroon van de twee groepen (r = .97, df = 28, p <.01 (gebaseerd op 6 x 5 matrix, inclusief kolom 'anders'). Er zijn slechts drie noemenswaardige (min of meer arbitrair op 8% of meer gesteld) verschillen. Ouderen herkennen Nederlandse erfwoorden en Franse leenwoorden beter dan jongeren en de vwo proefpersonen kennen relatief veel Franse leenwoorden toe aan het Engels. Bij de Engelse leenwoorden is er nauwelijks verschil tussen de twee groepen. Het percentage correct is vrijwel identiek (63% en 64%) en ook de misinterpretaties vertonen een vergelijkbaar patroon van toekenningen aan de verschillende talen. Ouderen zijn dus niet beter in staat de herkomst van Engelse leenwoorden te herkennen dan jongeren. In dit opzicht is er geen sprake van een leerproces.

Tabel 12 Percentages correcte identificatie (op de diagonaal) en verwarringen afzonderlijk voor de (jongere) vwo- en (oudere) post-vwo-proefpersonen (tussen haakjes). Met grijs is aangeduid dat er sprake is van een verschil van 8% of meer tussen de twee groepen proefpersonen.

		Responsie				
		Ned	Duits	Engels	Frans	Latijn
Stimulus	Ned	62 (75)	23 (18)	9 (4)	2 (1)	4 (2)
	Duits	55 (59)	28 (30)	6 (2)	3 (3)	6 (4)
	Engels	10 (8)	4 (3)	63 (64)	8 (12)	14 (12)
	Frans	20 (19)	7 (3)	22 (14)	24 (38)	26 (24)
	Latijn	21 (24)	9 (6)	21 (13)	13 (20)	35 (34)
	Totaal	36 (39)	15 (13)	21 (16)	10 (15)	17 (15)

3.3 Effect van het vak Latijn

In Tabel 13 zijn de responsies opgesplitst voor de personen die wel en hen die geen Latijn in hun vakkenpakket hebben (gehad). De proefpersonen met Latijn doen het over de hele linie ietsje beter (45.8%) dan de proefpersonen zonder Latijn (42.2%). Uit Tabel 13 kan worden opgemaakt dat er een grote overeenkomst is tussen de twee groepen proefpersonen in het responsiepatroon (r =.97, df = 28, p <.01, zie Sectie 3.2.2). Er zijn slechts twee noemenswaardige verschillen. De proefpersonen met Latijn herkennen Latijnse leenwoorden beter (dit is slechts voor een klein deel te wijten aan response bias) en zij zien relatief veel Franse woorden aan voor Latijnse leenwoorden.

Tabel 13 Percentages correcte identificatie (op de diagonaal) en verwarringen afzonderlijk voor de proefpersonen zonder Latijn en met Latijn (tussen haakjes). Met grijs is aangeduid dat er sprake is van een verschil van 8% of meer tussen de twee groepen proefpersonen.

		Responsie				
		Ned	Duits	Engels	Frans	Latijn
Stimulus	Ned	70 (68)	18 (23)	6 (6)	1 (1)	4 (1)
	Duits	59 (55)	26 (33)	4 (3)	3 (2)	6 (4)
	Engels	12 (6)	4 (2)	60 (67)	12 (8)	10 (16)
	Frans	22 (17)	5 (5)	18 (16)	34 (30)	20 (31)
	Latijn	25 (20)	8 (8)	18 (15)	19 (14)	28 (42)
	Totaal	40 (35)	13 (15)	18 (18)	14 (11)	14 (19)

Conclusie

In dit onderzoek hebben we ons een beeld proberen te vormen van de perceptie van 'gewone' taalgebruikers van de herkomst van gangbare woorden in het hedendaags gesproken Nederlands. Daartoe hebben we een schriftelijke enquête uitgevoerd waarin verschillende groepen informanten de herkomst moesten aangeven van in totaal 137 woorden. We waren daarbij geïnteresseerd zowel in de correcte als de incorrecte (afwijkend van de in de woordenboeken aangegeven) toekenningen. Onze hypothese was dat de oorsprong van Engelse leenwoorden makkelijker zou worden herkend dan de oorsprong van Franse, Duitse en Latijnse leenwoorden. Onze hypothese werd bevestigd. Voor niet-taalkundig geschoolden, onafhankelijk van leeftijd of kennis van het Latijn, zijn Engelse leenwoorden veel beter herkenbaar (64%) dan leenwoorden met een andere herkomst. Vooral Duitse leenwoorden worden slecht (29%) herkend en ook veel Franse en Latijnse woorden worden als Nederlandse erfwoorden of woorden met een andere herkomst gezien.

Dat Engelse leenwoorden zo goed herkenbaar zijn, kan worden verklaard uit verschillende factoren. Het feit dat veel Engelse woorden hun oorspronkelijke, van het Nederlands afwijkende spelling hebben gehandhaafd lijkt in belangrijke mate bij te dragen tot het gemak waarmee uit het Engels afkomstige woorden

worden herkend. Dit kwam naar voren uit een multipele regressie-analyse. De Engelse leenwoorden zijn in deze opzichten duidelijker als Engels gemarkeerd dan bijvoorbeeld de Duitse leenwoorden. Ook de grote algemene kennis van het Engels zal zeker een rol hebben gespeeld.

De grote herkenbaarheid van Engelse leenwoorden kan bijdragen aan het gangbare gevoel dat het Nederlands 'overspoeld' wordt door woorden uit het Engels. Hoewel het aantal woorden van Engelse herkomst op het totale hedendaagse lexicon gering is (in de huidige studie maken ze 1.5% uit van alle inhoudswoorden, Van der Sijs (1996, p. 64-67) komt uit op 2.3%), valt hun aanwezigheid door hun geringe mate van linguïstische aanpassing de hedendaagse taalgebruiker erg op. Bovendien zijn er concentraties van Engelse leenwoorden in een aantal populaire en in het oog vallende maatschappelijke domeinen, zoals de computer, personeelsadvertenties en reclame. In deze domeinen, en wellicht ook meer in het algemeen, levert het Engels het merendeel van nieuwe leenwoorden.

Bibliografie

Giesbers 2008 – C. Giesbers: *Dialecten op de grens van twee talen. Een dialectologisch en sociolinguistisch onderzoek in het Kleverlands dialectgebied*. Proefschrift Radboud Universiteit Nijmegen, Nijmegen, 2008.

Gooskens 1997 – C. Gooskens: *On the Role of Prosodic and Verbal Information in the Perception of Dutch and English Language Varieties.* Proefschrift Katholieke Universiteit Nijmegen, Nijmegen, 1997.

Preston 1999 – D.R. Preston: 'Introduction'. In: D.R. Preston (ed.): *Handbook of Perceptual Dialectology*, vol. 1. Amsterdam/Philadelphia, 1999, p. xxiii-xl.

Van Bezooijen & Heeringa 2006 – R. van Bezooijen & W. Heeringa: 'Intuitions on Linguistic Distance. Geographically or Linguistically Based?' In: *Artikelen van de Vijfde Sociolinguïstische Conferentie*. Delft, 2006, p. 77-87.

Van Bezooijen & Ytsma 2000 – R. van Bezooijen & J. Ytsma: 'Accents of Dutch. Personality Impression, Divergence and Identifiability'. In: R. Belemans & R. Vandekerckhove (eds.): *Variation in (Sub)Standard Language*. Amsterdam, 2000, p. 105-129.

Van Heuven, Neijt & Hijzelendoorn 1995 – V.J. van Heuven, A.H. Neijt & M. Hijzelendoorn: 'Automatische indeling van Nederlandse woorden op basis van etymologische filters'. *Spektator* 23 (1995), p.279-291.

Van Hout & Münstermann 1988 – R. van Hout & H. Münstermann: 'Linguïstische afstand, dialect en attitude'. *Gramma* 5 (1988), p. 101-123.

Van der Sijs 1996 – N. van der Sijs: *Leenwoordenboek. De invloed van andere talen op het Nederlands*. Den Haag/Antwerpen, 1996.

Van Veen & Van der Sijs 1997 – P.A.F. van Veen & N. van der Sijs: *Etymologisch woordenboek. De herkomst van onze woorden*. Utrecht/Antwerpen, 1997.

Weijnen 1966 – A. Weijnen: *Nederlandse Dialectkunde.* Assen, 1966.

Corrrespondentie-adres van de auteurs

Charlotte Gooskens
Afdeling Scandinavistiek RUG
Postbus 716
NL-9700 AS Groningen
c.s.gooskens@rug.nl

Bijlage Afgevraagde woorden in de enquête

Informeel

Nederlands	Duits	Engels	Frans	Latijn
aangenaam	eenzaam	bus	apart	april
boel	erts	club	cadeau	centrum
eigenlijk	kast	film	dialect	graad
glas	ogenblik	foto	fout	kaas
kwaad	omgeving	mail	krant	koken
missen	opname	shit	letter	museum
plek	trui	sport	oranje	pagina
spul	wedstrijd	team	pensioen	procent
verhaal	sowieso	video	raar	vakantie
wennen		weekend	simpel	vieren

Formeel

Nederlands	Duits	Engels	Frans	Latijn
behoren	bevoegd	infrastructuur	accepteren	actie
bieden	bewerkstelligen	partner	constateren	commissaris
gelegenheid	bewust	tanker	document	conferentie
immers	gevaar	internationaal	fraude	crisis
nauw	heersen	structureel	interventie	effect
overheid	invloed		liberaal	fungeren
stelsel	maatregel		paragraaf	orgaan
verdedigen	onderhavig		proces	regio
vertrouwen	overigens		stabiliteit	strikt
wapen	straf		versie	visie

Stijlneutraal

Nederlands	Duits	Engels	Frans	Latijn
ander	grens		activiteit	absoluut
betalen	herinneren		contact	basis
deur	ongeveer		enorm	collega
gaan	praktisch		idee	functie
hard	reageren		kans	kort
komen	regelmatig		modern	markt
misschien	richting		periode	proberen
schip	technisch	internet	persoon	project
spelen	toestand	plannen	principe	staat
verkeerd	verzamelen	starten	relatie	tafel