

Causale connectieven in het leerdercorpus Nederlands

Liesbeth Degand & Julien Perrez

*Université catholique de Louvain
Unité de Linguistique théorique et appliquée des Langues germaniques*

1. Inleiding

Dit artikel¹ wil een stand van zaken geven van de aanleg en opbouw van het LeerderCorpus Nederlands (LCN), een lopend project binnen de vakgroep Nederlandse Taalkunde van de Université catholique de Louvain. Om een beter inzicht te verkrijgen in het leerproces van een vreemde taal en de problemen die de leerder daarbij ondervindt, is het nodig de kenmerken van de tussentaal te bestuderen. Dergelijk onderzoek is het best gebaat met een degelijke empirische basis, zoals die voorhanden is in een corpus. De voordelen van (leerder)corpus analyses voor dit type onderzoek zijn tweevoudig: (i) het gaat om grote verzamelingen natuurlijke taaldata, die met behulp van specifieke automatische procedures onderzocht kunnen worden in functie van het gewenste standpunt (Granger 1998, Meunier 1998), en (ii) corpusonderzoek verschaft de tekstuele context waarin een gegeven woord of structuur verschijnt, zodat de semantische en syntactische context die het gebruik van deze structuren beregelen blootgelegd kan worden. Door deze L2-gegevens te contrasteren met vergelijkbare L1-data kunnen specifieke tussentaalverschijnselen aan het licht gebracht worden, zoals vermijdingsstrategieën, over-, onder-, of incorrect gebruik van de onderzochte items (Granger & Tyson 1996, Osborne 1998, Liu & Shaw 2001).

Leerdercorpusonderzoek heeft zijn deugdelijkheid al bewezen (Granger 1994, Granger, Hung & Petch-Tyson 2002), maar voor het Nederlands zijn de beperkingen duidelijk:

- Er bestaat geen leerdercorpus voor het Nederlands. Het moet dus worden samengesteld. Tot op heden hebben onderzoekers hun gegevens inderdaad op een ad-hoc basis verzameld, in functie van de doeleinden van hun eigen onderzoek, wat hergebruik en distributie in sterke mate beperkt.
- Dergelijke ad-hoc "corpora" zijn bovendien beperkt qua omvang. Ze zijn dus wel nuttig voor de studie van morfo-syntactische verschijnselen, maar voor de studie van lexicale

en/of discursieve aspecten voldoen ze niet. Een eerste doelstreven voor het LCN is bijgevolg 1.000.000 woorden.

- Franstaligen vormen de grootste en meest homogene groep leerders van het Nederlands. Bij de opbouw van het LCN willen we daarom uitgaan van deze groep, waarbij we ons in een eerste fase willen beperken tot geschreven taal. Uitbreidingen naar andere groepen leerders en naar gesproken taal kunnen plaatsvinden in latere stadia van het project.

2. Opbouw van het LCN

Het hoofddoel van dit project is dus de verzameling en opbouw van een LeerderCorpus voor het Nederlands (Perrez & Degand 2003). Daarbij zullen we in de grote lijnen het ICLE project volgen (Granger 1993), een internationaal leerdercorpus van het Engels, dat bestaat uit een verzameling (argumentatieve en literaire) opstellen van 500 tot 1000 woorden, geschreven door derde- en vierdejaarsstudenten van 19 verschillende moedertaalachtergronden. Elk subcorpus (per moedertaal) bestaat uit 200.000 woorden. Voor elk opstel wordt er met behulp van een aantal variabelen een leerdersprofiel en schrijftaakprofiel bijgehouden. Die moeten het mogelijk maken niet alleen kwantitatieve gegevens te verzamelen over het gebruik van bepaalde taalstructuren, maar vooral ook kwalitatieve analyses uit te voeren waarbij hypothesen i.v.m. leeftijd, studieniveau, opleidingstype e.d.m. getoetst kunnen worden. Bij de aanleg van het LCN wijken we op een aantal punten af van de ICLE-lijn. Zoals reeds vermeld willen we ons, althans in de beginfase van het project, beperken tot leerders van het Nederlands die het Frans als moedertaal hebben. Zij vormen immers de grootste en meest homogene groep Nederlands-als-vreemde-taal-lerenden: meer dan 70% van de Franstalige Belgen leert tijdens een of andere periode in zijn schoolcurriculum Nederlands. Onderzoeksresultaten en pedagogische implicaties zullen dus naar alle waarschijnlijkheid van belang zijn voor een groot aantal didactici. Ten tweede willen we meer variatie toelaten voor de leerdervariabelen "studieniveau" en "opleiding", door niet alleen schrijfp opdrachten te verzamelen van de meest gevorderde "letteren" studenten (derde- en vierdejaars), maar ook laatstejaars uit de middelbare school en eerste- en tweedejaarsstudenten uit gevarieerde opleidingen (letteren, economie, rechten, ...). Die variatie in studieniveau is noodzakelijk als we willen onderzoeken of bepaalde tussentaalverschijnselen (in het gebruik van discourse markers bijvoorbeeld) op een of andere manier gerelateerd kunnen worden aan de ontwikkeling van de taalvaardigheid. De andere leerdervariabelen betreffen de "leeftijd" en het "geslacht", de "streek van herkomst", de

"kennis van andere talen" en het "aantal jaren Nederlands" dat de leerder achter de rug heeft. Ook de schrijftaakvariabelen willen we onder controle houden, zoals het "genre" (opstel, brief, schrijfoefening, ...), en de "schrijfcontext" (in de klas, thuis, als huiswerk, met of zonder referentiewerken).

Om de typische tussentaalverschijnselen te bestuderen moet echter eveneens een "controlecorpus" aangelegd worden, d.i. een vergelijkbare verzameling data van moedertaalschrijvers. We beschikken al wel over talrijke krantencorpora, maar die zijn qua genre, leeftijd van de schrijvers en schrijfcontext niet vergelijkbaar met het (toekomstige) LCN. De aanleg van dit controlecorpus (of NCN – Native Corpus Nederlands) zal parallel verlopen met het LCN.

Tenslotte is dit type onderzoek ook best gebaat met een getagged corpus, zodat specifieke structuren op een efficiënte wijze kunnen worden opgezocht. We denken hierbij aan POS-tagging (Part-of-Speech) (Van Eynde, Zavrel & Daelemans 2000) en error-tagging (spelling, maar ook grammaticale en fraseologische fouten) (Dagneaux, Denness & Granger 1998). Dit zal een nauwkeurige beschrijving van het Nederlands als tussentaal van Franstaligen bevorderen. Het POS-taggen zelf kan semi-automatisch gebeuren, maar tot nog toe is error-tagging in de hoofdzaak manueel. Dit verklaart waarom zo'n corpus totnogtoe niet voorhanden is voor het Nederlands, terwijl het wel enorme onderzoeksperspectieven biedt.

3. LCN: Stand van zaken

Voor de aanleg van het LCN zijn we in de eerste plaats uitgegaan van bestaand materiaal dat we voor onze doeleinden hebben "gerekupereerd". We hebben dus contact opgenomen met collega's aan de andere Franstalige universiteiten in België, met de organisatoren van de ANBF-opstellenwedstrijd, en met het Certificaat Nederlands als Vreemde Taal (CNAVt). Het resultaat van deze eerste ronde was succesvol, daar we onmiddellijk konden beschikken over opstellen van de eerste kandidatuur tot en met de tweede licentie uit Luik (ULg), Brussel (ULB) en Louvain-la-Neuve (UCL), alsook over opstellen van leerlingen van het zesde middelbaar uit de omgeving van Luik (materiaal uit Hiligsmann 1993) en de schrijfexamens uitgebreide kennis van het CNAVt (examenjaar 2002). Deze data waren grotendeels niet elektronisch beschikbaar en werden dus ingetikt en ingevoerd in een database. Tabel 1 geeft de huidige (kwantitatieve) stand van zaken.

Tabel 1: Samenstelling van het LCN

Studieniveau	Aantal woorden	Aantal opstellen
	260.179	820
6 ^{de} middelbaar	60.453	327
1 ^{ste} kandidatuur/regentaat	98.127	360
2de kandidatuur/regentaat	53.976	101
1 ^{ste} licentie/3 ^{de} regentaat	6865	15
2 ^{de} licentie	9978	12

In de komende maanden (en waarschijnlijk jaren) willen we het corpus zo uitbreiden dat we ons streefdoel van één miljoen woorden bereiken. De verzameling van dit nieuwe materiaal moet op een systematische manier gebeuren, om te vermijden dat het corpus te eenzijdig wordt en niet meer representatief zou zijn. Hierbij denken we vooral aan de verschillende studieniveaus (de laatste jaren zijn in onze voorlopige data ondergerepresenteerd) en studierichtingen (totnogtoe hebben we bijna uitsluitend producties van studenten uit de studierichting germaanse talen).

Maar ook uit dit tamelijk beperkte corpus kunnen interessante onderzoeksgegevens over de tussentaal van Franstalige leerders van het Nederlands de kop opsteken. Dit willen we in de volgende paragrafen illustreren.

4. Distributie van discourse markeringen in NT-2

In het kader van lopend onderzoek naar de impact van connectieven op tekstbegrip in L1 en L2 zijn we tot de bevinding gekomen dat lezers van expositorische teksten baat hebben van de aanwezigheid van connectieven in een tekst (Degand, Lefèvre & Bestgen 1999, Degand & Sanders 2001). Concreet betekent dit dat lezers van korte vulgariserende wetenschappelijke teksten (ongeveer 250 woorden) minder moeite hebben met het beantwoorden van vragen over deze tekst wanneer een aantal causale relaties hierin expliciet zijn gemaakt door middel van causale connectieven dan wanneer deze relaties impliciet (ongemarkeerd) zijn gelaten. Een van de nevenresultaten van dit onderzoek was de bevinding dat de taalvaardigheid van de lezers een rol zou kunnen spelen op de impact van die connectieven op het tekstbegrip. Er is met name gebleken dat lezers met een (te) lage taalvaardigheid in een vreemde taal geen baat

(meer) hebben van de aanwezigheid van discourse markeringen. Verder onderzoek moet uitsluitsel geven over de voorwaarden van een dergelijke impact op het tekstbegrip van tweede-taalleerders (Perrez & Degand, in voorbereiding).

In het verlengde van dit onderzoek naar de interactie tussen aan- of afwezigheid van (causale) connectieven en tekstbegrip in een vreemde taal willen we onderzoeken hoe het zit met de tekstproductie van vreemde-taalleerders. Is het zo dat een adequaat gebruik van discourse markeringen in tekstproductie hand in hand gaat met een geschikte interpretatie in tekstbegrip (zonder begrip geen productie), of is het in productie veeleer de rol van de schrijfvaardigheid in het algemeen, en met name in de moedertaal, die een doorslaggevende rol zal spelen? Volgens deze laatste stelling zouden goede moedertaal-schrijvers “automatisch” ook goede vreemdetaal-schrijvers zijn doordat ze hun schrijfprocesvaardigheden zouden transfereren naar de vreemde taal (Cumming 1989, Lefrancois 2001, Schoonen, Van Gelderen, De Glopper, Hulstijn, Simis, Snellings & Stevenson 2003, Whalen & Menard 1995). Volgens deze hypothese is (goede) schrijfvaardigheid in de hoofdzaak een kwestie van (cognitieve) ontwikkeling (Bereiter & Scardamalia 1987). Maar ook de rol van een eventuele talige transfer van de moedertaal naar de vreemde taal in het gebruik van discourse markeringen moet onder de loep genomen worden (Gass 1996). In dit laatste geval zouden de schrijvers hun kennis over het gebruik van discourse markeringen in hun moedertaal vertalen naar de “tegenhangers” in de vreemde taal. In de producties van vreemde taalschrijvers zou de invloed van de moedertaal aantoonbaar moeten zijn. In de volgende paragraaf willen we deze gedachtegang verder onderzoeken en illustreren aan de hand van de data die nu reeds voorhanden zijn in het LCN.

4.1. Het gebruik van causale connectieven in NT-2

Over de distributie en semantische kenmerken van Nederlandse causale connectieven zoals *dus* en *daarom* of *want* en *omdat* is in de taalkundige literatuur al heel wat verschenen (voor een overzicht, zie Degand 2001, Pit 2003). Algemeen wordt aangenomen dat causale connectieven expliciete markeringen zijn van een onderliggende causale coherentierelatie. Met de hulp van een causaal connectief wordt de causale relatie tussen twee tekstsegmenten expliciet gemaakt, in die zin heeft men het ook over relationele markeringen. De keuze van het connectief is afhankelijk van de causale situatie die uitgedrukt moet worden. Zo heeft het Nederlands bijvoorbeeld de connectieven *doordat* en *daardoor* die gespecialiseerd zijn in het uitdrukken van zogenaamd *non-volitionele causale* relaties (zie voorbeelden (1-2)), terwijl

andere connectieven zoals *want*, *daarom* of *dus* dan weer geschikter zijn om andere soorten relaties uit te drukken, zoals *volitionele* relaties (3), *epistemische* (4) of *interactionele* (5) (zie verder Degand & Pander Maat 2003; Pander Maat & Degand 2001, Pit, Pander Maat & Sanders 1997).

- (1) De jongen raakte gewond **doordat** hij niet ontploft vuurwerk aanstak.
- (2) Het had de hele nacht gestormd. **Daardoor** waren er verschillende pannen van het dak gevallen.
- (3) Ze was dol op de sneeuw en was **daarom** extra vroeg op pad getrokken.
- (4) Ik heb al gespeeld, **dus** het is nu aan jou.
- (5) Kleed jullie nu snel aan, **want** we moeten zo weg.

Over hoe het zit met het gebruik van dergelijke connectieven in een vreemde taal is nog heel weinig bekend. In de literatuur over tweedetaalverwerving wordt gesuggereerd dat leerders van een vreemde taal geneigd zijn tot overgebruik en incorrect gebruik van discourse markerings. Granger & Tyson (1996) en Osborne (1999) wijzen er bijvoorbeeld op dat Franstalige leerders van het Engels een overgebruik tonen van connectieven zoals *indeed*, *moreover*, *of course*, *actually*, enz. Dit zijn connectieven die de spreker helpen zijn discourse te oriënteren en te motiveren (Osborne 1999: 100). Het gaat met andere woorden om *metadiscursieve* markerings die de attitude van de spreker ten opzichte van zijn discourse uitdrukken door bijvoorbeeld aan te geven hoe de spreker zijn argumenten structureert en motiveert (Hyland 1998, Intaraprawat & Steffensen 1995). Hoe zulke markerings door Franstalige leerders van het Nederlands gebruikt worden, wordt elders beschreven (Perrez 2004). Hier willen we ons verder concentreren op relationele markerings die een causaal verband leggen tussen twee tekstsegmenten. Op het eerste gezicht zou hier geen sprake moeten zijn van overgebruik, er is immers geen vanzelfsprekende reden om aan te nemen dan leerders vaker causale relaties zouden markeren dan moedertaal-schrijvers.

Tabel 2 geeft de distributie in het leerdercorpus van de in NT-1 meest frequente causale connectieven (Degand 2001). De frequentie van de markerings in het leerdercorpus werd vergeleken met hun frequentie in een controlecorpus, samengesteld uit een verzameling artikelen uit 'De Limburger'.

Tabel 2: Causale connectieven in het leerdercorpus vs. NT1-krantencorpus

Connectieven	LCN	De Limburger
	204.751 woorden	901.916 woorden
	frequentie/100.000	frequentie/100.000
Omdat	355	103
Want	170	55
Aangezien	12	4
Doordat	1	6
Dus	303	47
Daarom	75	33
Daardoor	12	14
Totaal	1028	262

Wat in Tabel 2 meteen opvalt, is dat de causale connectieven wel massaal worden overgebruikt door de Franstalige schrijvers, met name *omdat* en *want* komen drie keer vaker voor in het LCN dan in de Limburger, terwijl *dus*² tot zes keer vaker voorkomt! Het lijkt er dus op dat de causale relaties door Nederlandstalige schrijvers (journalisten!) vaker impliciet worden gelaten. Deze resultaten moeten echter meteen worden genuanceerd. Het is immers niet onproblematisch om argumentatieve opstellen in een tweede taal zomaar te vergelijken met krantenartikelen. Het gaat namelijk om verschillende genres waarin het gebruik van connectieven en coherentierelaties sterk kan variëren (Biber 1988, Sanders 1997). Om hieraan tegemoet te komen, zouden de NT2-opstellen moeten worden vergeleken met NT1-opstellen. De verzameling van een dergelijk L1-corpus is lopende en is onontbeerlijk om de discursieve kenmerken van de tussentaal van de leerders nauwkeurig te kunnen bepalen. Dit neemt niet weg dat het overgebruik in L2 dermate hoog ligt, dat een verklaring niet exclusief gezocht mag worden in genreverschillen. Wat we hier grondiger willen bekijken is of de moedertaal van de L2-schrijvers hier een invloed heeft gehad, dan wel of er andere factoren een rol zouden kunnen spelen.

4.2. Overgebruik van markeringsen: L1-Transfer of niet?

Granger & Tyson (1996) relateren het overgebruik van hun onderzochte connectieven aan een transfereffect uit de moedertaal van de leerders. In overeenstemming met Hervey & Higgins (1992) die stellen dat “it is more common in French than in English for texts to be explicitly structured by the use of connectives” (Hervey & Higgins 1992: 49) gaan Granger & Tyson (1996) ervan uit dat het Franse retorische systeem meer gebruik maakt van expliciterende talige uitdrukkingen zoals connectieven en dat de leerders derhalve dat retorische systeem toepassen op hun tweede taal. Dit resulteert in soms grote verschillen tussen de distributie van bepaalde connectieven in de moedertaal van de leerders en hun ‘equivalenten’ in de doeltaal. Om hun transferstelling te onderbouwen, vergelijken Granger & Tyson (1996) producties van leerders met verschillende moedertaalachtergronden en stellen zo vast dat Franstalige leerders een connectief als *indeed* vaker gebruiken dan Duitstalige leerders. Op basis hiervan concluderen ze dat deze tendens tot overgebruik te maken heeft met een transfereffect uit het Frans.

Een andere manier om de transferhypothese te toetsen bestaat in de vergelijking van de NT2-producties van de Franstalige leerders met T1-producties van Franstalige schrijvers (zie ook Perrez 2004). De vraag is of het gebruik van causale connectieven ook zo omvangrijk is als Franstalige schrijvers opstellen schrijven in hun moedertaal. Tabel 3 illustreert de frequentie van de onderzochte markerings respectievelijk in het leerdercorpus en in een corpus bestaande uit opstellen die door Franstalige schrijvers zijn geschreven in hun moedertaal³. Hierbij moet worden opgemerkt dat de equivalentie tussen Franse en Nederlandse uitdrukkingen zoals *dus/donc*, *want/car*, *omdat/parce que* ... hypothetisch is. We beweren niet dat deze uitdrukkingen even vaak voorkomen in beide talen en dat ze hetzelfde betekenen in alle contexten. De Franse equivalenten worden vooral ter illustratie aangeboden.

Tabel 3: Causale connectieven in het leerdercorpus vs. FT1-opstellen

Causale connectieven	LCN (freq./100.000)	Causale connectieven	Opstellen Frans (freq.:100.000)
Omdat	355	parce que	64
Want	170	car	129
aangezien	12	puisque	42
Doordat		-	-
Dus	303	donc	172
Daarom	75	c'est pourquoi	18

Wat in Tabel 3 in het oog springt, is dat het overgebruik van *omdat*, want en *dus* door de Franstalige leerders niet onmiddellijk te wijten kan zijn aan het gebruik van hun "equivalenten" in de moedertaal. De L1-transfer hypothese dus blijkbaar niet zomaar gevalideerd worden. Het lijkt er veeleer op dat de leerders in de eerste plaats een eentonig gebruik maken van causale connectieven in hun schrijfproducten, met als koploper *omdat*. Hierbij wordt voorbij gegaan aan meer specifieke causale connectieven zoals *doordat*⁴, *daardoor* en *aangezien*. Een vluchtige analyse van Tabel 4, waarin het gebruik van de connectieven in functie van het leerjaar van de leerders wordt gegeven, toont aan dat er in dit opzicht een evolutie aan de gang is: naarmate de leerders bekwamer worden in hun tweede taal, maken ze meer gebruik van de andere mogelijkheden om de causale relaties expliciet te maken. Terwijl het aantal voorkomens van *omdat* en *want* afneemt, neemt het aantal voorkomens van *aangezien*, *daardoor* en (in mindere mate) *doordat* toe. Deze dubbele beweging stemt overeen met de observatie van Kerr-Barnes (1998: 190) dat

“basic patterns of acquisition are similar in that younger/less advanced subjects show greater use of high-frequency connectors while older/more advanced subjects show an expanded repertoire of connectors, together with increased differentiation of function, as well as signalling of more complex semantic and pragmatic relations.”

Of deze evolutie inderdaad gepaard gaat met een breder scala aan soorten causale relaties, zoals Kerr-Barnes beweert, zal een kwalitatieve analyse moeten uitmaken.

Tabel 4: Causale connectieven in LCN. Vergelijking tussen drie leerjaren.

Connectieven	LCN Niv.1	LCN Niv.2	LCN Niv.3
	60.453 w.	93.371 w.	50.927 w.
	freq./100.000	freq./100.000	freq./100.000
Omdat	356	495	259
Want	248	144	124
Aangezien	8	8	23
Doordat	0	0	4
Dus	371	342	235
Daarom	76	151	129
Daardoor	5	16	29

Samenvattend kunnen we dus stellen dat schrijfvaardigheid in een vreemde taal, zowel als in de moedertaal, het resultaat is van een evolutionair proces (Bereiter & Scardamalia 1987). Dit neemt echter niet weg dat de moedertaal ook een invloed kan hebben op het gebruik van bepaalde talige structuren, met name het ondergebruik en incorrect gebruik vergt aanvullende kwalitatieve analyses.

5. Conclusie

In deze bijdrage hebben we een overzicht gegeven van de structuur en opbouw van het LeerderCorpus Nederlands (LCN), een lopend onderzoek dat de aanleg van een corpus van schrijfproducten van Franstalige leerders van het Nederlands beoogt. De uiteindelijke doelstelling van het project is te beschikken over een corpus van één miljoen woorden verdeeld over vier à vijf schrijfvaardigheidsniveaus.

Hoewel het corpus op dit ogenblik nog tamelijk beperkt is, hebben we getracht aan te tonen dat het nu reeds een hoogst interessante bron vormt voor het onderzoek van de tussentaal van Franstaligen die Nederlands leren. Met name de analyse van het gebruik van causale connectieven wijst op een overgebruik én eentonig gebruik van causale markeringen. Dit overgebruik lijkt niet te wijten te zijn aan een transfer uit de moedertaal (Frans) naar de vreemde taal (Nederlands). Een verklaring moet veeleer gezocht worden in de evolutie die elke vreemde-taalleerder doormaakt in zijn schrijfvaardigheid. Naarmate de leerders de vreemde taal beter beheersen, treedt ook meer variatie op in de keuze van de causale markeringen. Aanvullende kwalitatieve analyses zouden moeten uitmaken in hoever deze eerste bevindingen bevestigd kunnen worden.

6. Referenties

BEREITER C. & M. SCARDAMALIA (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

BIBER, D. (1988). *Variation across Speech and Writing*. London: Cambridge University Press.

CUMMING, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39, 81-141.

DAGNEAUX, E., DENNESS, S. & S.GRANGER (1998). Computer-aided Error Analysis. *System. An International Journal of Educational Technology and Applied Linguistics*, 26, 163-174.

DEGAND, L. (2001). *Form and Function of Causation. A Theoretical and Empirical Investigation of Causal Constructions in Dutch*. Leuven: Peeters.

DEGAND, L., LEFEVRE, N. & Y. BESTGEN (1999). The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design*, 1, 39-51.

DEGAND, L. & H. PANDER MAAT (2003). Contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In: VERHAGEN, A. & J. VAN DE WEIJER (eds.) *Usage based approaches to Dutch*. Utrecht: LOT, 175-199.

DEGAND, L. & T. SANDERS (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, 15, 739-758.

ELLIS, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

GASS, S. (1996). Second language acquisition and linguistic theory: The role of language transfer. In: RITCHIE, W.C. (ed.) *Handbook of Second Language Acquisition*, San Diego, Academic Press, 317-345.

GRANGER, S. (1993). The International Corpus of Learner English. In: AARTS, J., DE HAAN, P. & N. OOSTDIJK (eds). *English Language Corpora: Design, Analysis and Exploitation*. Rodopi, Amsterdam, 57-69.

GRANGER, S. (1994). The Learner Corpus: A Revolution in Applied Linguistics. *English Today*, 39, 25-29.

GRANGER, S. (ed.) (1998). *Learner English on Computer*. Addison Wesley Longman, London & New York.

GRANGER, S., DAGNEAUX, E. & F. MEUNIER (2002). *The International Corpus of Learner English. Handbook and CD-ROM*. Presses Universitaires de Louvain: Louvain-la-Neuve.

- GRANGER, S., HUNG, J. & S. PETCH-TYSON (eds) (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Benjamins, Amsterdam and Philadelphia.
- GRANGER, S. & S. TYSON (1996). Connector usage in the English essay-writing of native and non-native EFL speakers of English. *World Englishes*, 15, 17-27.
- HERVEY, S. & I. HIGGINS (1992). *Thinking Translation*. London: Routledge.
- HILIGSMANN, P. (1993). *Onderzoek naar enkele aspecten van de tussentaal van Franstalige M.O.-leerders van het Nederlands. Linguïstische en didactische aspecten*. Doctorale dissertatie : Université de Liège, 2002-2003.
- HYLAND, K. (1998). Persuasion and context: the pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30, 437-455.
- INTARAPRAWAT, P. & M.S. STEFFENSEN (1995). The use of metadiscourse in good and poor ESL essays. *Journal of Second Language Writing*, 4, 253-272.
- KERR-BARNES, B. (1998). The acquisition of connectors in French L2 narrative discourse. *Journal of French Language Studies*, 8, 189-208.
- LEFRANÇOIS, P. (2001). The issue of transfers in second-language texts. *Canadian modern language review – Revue canadienne des langues vivantes*, 58, 223-245.
- LIU, E.T.K. & P.M. SHAW (2001). Investigating learner vocabulary: a possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL*, 39, 171-194.
- MEUNIER, F. (1998). Computer tools for the analysis of learner corpora. In: GRANGER, S. (ed.), *Learner English on computer*, New-York, Addison Wesley Longman, 19-37.
- OSBORNE, J. (1999). L'utilisation des marqueurs de liaison en Anglais: étude inter-culturelle des productions d'étudiants. *Actes des journées d'étude du GEPED*, Université de Paris VII, 97-109.
- PANDER MAAT, H. & L. DEGAND (2001). Scaling causal relations and connectives in terms of Speaker Involvement. *Cognitive Linguistics*, 12, 211-245.
- PERREZ, J. (2004). Het gebruik van connectieven door Franstalige leerders van het Nederlands. *Toegepaste Taalwetenschap in Artikelen* 71.
- PERREZ, J. & L. DEGAND (2003). On the combination of corpus-based and experimental methodologies in the study of causal, contrastive and metadiscourse connectives in L1 and L2

text comprehension and production. *Proceedings of the Corpus Linguistics 2003 Conference*, Lancaster, 627-636.

PERREZ, J. & L. DEGAND (in voorbereiding). De impact van causale connectieven op het begrip van informatieve teksten door Franstalige leerders van het Nederlands. Voorgelegde paper op het colloquium 'Neerlandistiek in Frankrijk en Franstalig België', UCL, 15-17 maart 2004.

PIT, M. (2003). *How to express yourself with a causal connective: subjectivity and causal connectives in Dutch, German and French*. Doctorale dissertatie. Rodopi: Amsterdam.

PIT, M., PANDER MAAT, H., & T. SANDERS (1997). 'Doordat', 'omdat', 'want'. Perspectieven op hun gebruik. *Taalbeheersing*, 19, 238-251.

SANDERS, T. (1997). Semantic and pragmatic sources of coherence: on the categorization of coherence relations in context. *Discourse Processes*, 24, 119-148.

SCHOONEN, R., VAN GELDEREN, A., DE GLOPPER, K., HULSTIJN, J., SIMIS, A., SNEELLINGS, P. & M. STEVENSON (2003). First language and second language writing: The role of linguistic knowledge, speed of processing and metacognitive knowledge. *Language Learning*, 53, 165-202.

VAN EYNDE, F., ZAVREL, J. & W. DAELEMANS (2000). Part of Speech Tagging and Lemmatisation for the Spoken Dutch Corpus. In: GRAVILIDOU, M., CARAYANNIS, G., MARKANTONATOU, S., PIPERIDIS, S. & G. STAINHAOUER (eds.), *LREC-2000 (Second International Conference on Language Resources and Evaluation) Proceedings*.

WHALEN K. & N. MENARD (1995). L1 and L2 writers strategic and linguistic knowledge – A model of multiple-level discourse processing. *Language Learning*, 45, 381-418.

¹ Liesbeth Degand is onderzoeker aan het Fonds National de la Recherche Scientifique. Dit project wordt ondersteund door projecten n° 03/08-301 van de Communauté française de Belgique – Actions de recherche concertées en n° 2.4535.02 (Fonds de la Recherche Fondamentale Collective).

² Over het overgebruik van *dus* in leerderproducties, zie Perrez (2004).

³ Het French Academic Essay Corpus werd samengesteld door het Centre for English Corpus Linguistics, UCL en bestaat uit opstellen geschreven door Franstalige studenten uit de studierichting romaanse aan de UCL. Het corpus bevat 93.555 woorden.

⁴ Hierbij moet worden opgemerkt dat *doordat* geen echte tegenhanger heeft in het Frans, wat een intuïtief gebruik in de weg staat (zie ook Perrez 2004).