

DE BOER, Minne Gerben. 'Selectie van Italiaanse werkwoordsvormen op basis van frequentieonderzoek'. Hoofdstuk 1 van *Woordstudies I*. ITALIANISTICA ULTRAIECTINA 4. Utrecht: Igitur, Utrecht Publishing & Archiving Services, 2009. ISBN 978-90-6701-025-2. 7-17.

SAMENVATTING

Dit artikel is een vroege recensie (mei 1972) van het LIF, de frequentielijst van IBM Italia uit 1971. De auteur concentreert zich op de verschillen met het project van Juilland, dat als uitgangspunt voor het LIF heeft gefungeerd. Op grond van de gegevens van het LIF is een basislijst van 50 werkwoorden opgesteld; verder wordt er een overzicht gegeven van de belangrijkste onregelmatige werkwoordsvormen.

RIASSUNTO

Questo articolo è una delle prime recensioni (maggio 1972) del LIF, la lista di frequenza di IBM Italia del 1971. L'autore si concentra sulle differenze rispetto al progetto di Juilland, che fu il modello del LIF. In base ai dati di questo lessico è stato compilato un elenco di base di 50 verbi italiani nonché una scelta delle principali forme verbali irregolari.

SUMMARY

This article is an early review (May 1972) of LIF, the frequency list of Italian words established by IBM Italia in 1971. The author concentrates on the differences with respect to the Juilland project, which had served as a model for LIF. On the base of the LIF data a basic list of 50 Italian verbs has been established, as well as an overview of the principal irregular verb forms.

KEY-WORDS

Frequency research, basic word list, irregular verb forms

© Minne G. de Boer

De bundel *Woordstudies I* vormt het vierde deel van de reeks ITALIANISTICA ULTRAIECTINA. STUDIES IN ITALIAN LANGUAGE AND CULTURE, uitgegeven door Igitur, Utrecht Publishing & Archiving Services, ISSN 1874-9577 (<http://www.italianisticaultraiectina.org>).

1. SELECTIE VAN ITALIAANSE WERKWOORDSVORMEN OP BASIS VAN FREQUENTIEONDERZOEK*

Het kernbegrip bij de didactiek van het vreemde-talenonderwijs is *selectie*. Selectie, niet in de zin van het selecteren van leerlingen, maar in de zin van het selecteren van leerstof, dat wil zeggen van woorden, van paradigmata, van syntactische constructies en van leestukjes. De tijd van de leerling is beperkt, zijn geheugen heeft een beperkte capaciteit, en de potentiële leerstof is enorm groot. Een niet te omvangrijk woordenboek bevat al gauw een 50.000 woorden; een cursus daarentegen die erin slaagt om de leerling 3.000 woorden bij te brengen vereist al behoorlijk wat tijd en energie. Het is daarom uiterst belangrijk dat die 3.000 (of 2.000, of 4.000) woorden zo gekozen worden dat de leerling er een maximaal profijt van heeft. Een voorbeeld kan dit verduidelijken. De *Piccolo Garzanti* bevat volgens mijn gegevens 5.249 werkwoorden. Hierbij, zoals bij alles wat ik vandaag zal zeggen, moet ik onmiddellijk de reserve maken dat er telfouten kunnen zijn opgetreden; ieder getal is dus slechts bij benadering juist. Nemen wij hieruit een willekeurige keuze, bijvoorbeeld de eerste twintig werkwoorden, in alfabetische volgorde, die beginnen met de letter *b*. Het zijn: *bacare, bacchettare, bacchiare, baciare, badare, bagnare, balbettare, balbutire, balenare, balestrare, ballare, ballonzolare, balocarsi, baluginare, balzare, balzellare, bamboleggiare, bandire, barattare, barbugliare*. Bij het aanhoren van deze lijst zal het een ieder duidelijk zijn dat kennis van de meeste van deze werkwoorden een luxe is voor de buitenlander die Italiaans leert. Waarschijnlijk zal er ook een grote mate van overeenstemming zijn over de werkwoorden die wel tot de basiswoordenschat moeten behoren, te weten *baciare, badare, bagnare* en *ballare*. Ik vermoed dat er enige aarzeling zal zijn ten aanzien van *balbettare, balzare* en misschien *bandire*. Nu is het ongetwijfeld mogelijk met een aantal vakmensen bijeen te komen en door middel van een soort ballottage een communis opinio te bereiken over de op te nemen werkwoorden, waarbij eventueel in het bereikte aantal in een tweede ronde geschrapt wordt tot men bijvoorbeeld tot 1000 werkwoorden is aangekomen. Dit is de werkwijze die door Georges Matoré is gevolgd in zijn *Dictionnaire du Vocabulaire Essentiel*, waar een équipe van vijf mensen gekomen is tot een woordenschat van *Les 5000 mots fondamentaux*. Ten aanzien van de equivalenten van onze juist genoemde woorden kan ik melden dat *embrasser, mouiller* en *danser* inderdaad zijn opgenomen (*badare* heeft uiteraard geen direct equivalent). Voorts zijn opgenomen *sursauter* en *bondir* (equivalenten van *balzare*), maar noch *bégayer*, noch *bredouiller*, noch *balbutier*. Het woor-

* Lezing gehouden op de Didactische Dag van de sectie Italiaans van Levende Talen te Nijmegen op 22 mei 1972.

denboek van Matoré is ontstaan uit een polemische instelling tegenover frequentieonderzoek; ik citeer uit het voorwoord:

En l'absence de dépouillements exhaustifs, il nous a semblé préférable de nous référer à la conscience linguistique de Français cultivés et à l'expérience pédagogique que nous-mêmes, professeurs de français, avons pu acquérir dans la pratique de l'enseignement.

Hoewel de ervaring van de docent natuurlijk belangrijk is, lijkt mij dit criterium echter nogal gevaarlijk, in de eerste plaats vanwege een onvermijdelijke beroepsdeformatie van de docent (immers de meeste taaldocenten hebben tegelijk een literaire opleiding en hun *coscienza linguistica* is daardoor automatisch litterair getint); in de tweede plaats omdat op deze wijze gemakkelijk normatieve elementen in het oordeel kunnen meespelen. Een voorbeeld: alle Italiaanse schoolboekjes bevatten oefeningen waarin het werkwoord *fare* vervangen moet worden door een preciezer werkwoord. Ik citeer "*fare e trovare: ecco due verbi di gran giro nella lingua dei moderni scansafatiche*" (Pestelli, 1962, p. 157). Nu is het echter juist de taak van het taalonderwijs om van de leerlingen *scansafatiche* 'luiwammesen' te maken, in die zin dat ze zich zullen moeten leren behelpen met minder woorden dan ze zouden willen gebruiken als ze hun eigen taal spreken; de taaldidactische eis is daarom het maximum te halen uit dat ene werkwoord *fare*. Het criterium bij deze intuïtieve selectie is dus onzeker, en de twijfel wordt groter naarmate men de grens raakt van het aantal woorden dat men wil opnemen. Toch is dit tot voor kort het enige criterium geweest dat voor het Italiaans kon worden toegepast, gezien de achtergebleven stand van het Italiaanse taalonderzoek. Het resultaat is dan ook duidelijk voor wie de woordregisters van enkele methodes vergelijkt; zelden stemt de woordenschat overeen. Het afgelopen jaar heb ik gewerkt met een van de modernste cursussen, *Insieme a Venezia*, die in velerlei opzicht goed is opgezet, speciaal voor Fransen in ieder geval; wat mij hierbij het meest getroffen heeft is de enorme overvloed van relatief onbelangrijke woorden in de leesstukjes.

Het alternatief, dat overigens voor de meeste levende talen heel goed bekend is, is een selectie gebaseerd op frequentie in een representatief corpus. Daarom kunnen wij nu dan ook zeer blij zijn met het recente *Lessico di Frequenza della lingua Italiana Contemporanea* van Bortolini, Tagliavini en Zampolli, uitgegeven door de IBM Italia. Ondanks de bezwaren die aan deze lijst kleven, is ze nu de onmisbare basis voor iedere selectie voor het onderwijs in het Italiaans. Voordat ik nu verder ga met de presentatie van deze frequentielijst en de resultaten die ik eruit gehaald heb, wil ik even enkele corrigerende opmerkingen maken. Frequentie is belangrijk, maar niet het enige criterium voor selectie. Er zijn een aantal andere criteria die ik even kort wil opsommen, met een paar voorbeelden (ik baseer me voor deze criteria op Mackey 1969). De lessituatie is belangrijk: iemand die met een directe methode werkt, kan niet beginnen zonder de woorden *lavagna* 'schoolbord' en *gessetto* 'krijtje', hoewel het eerste maar driemaal voorkomt in de IBM-lijst en het tweede helemaal niet. Bepaalde

klassen van woorden moeten tezamen aangeboden worden ongeacht de relatieve frequentie, bijvoorbeeld de telwoorden of de namen der maanden. De IBM-lijst geeft: *gennaio* 14, *febbraio* 14, *marzo* 30, *aprile* 25, *maggio* 45, *giugno* 43, *luglio* 29, *agosto* 30, *settembre* 24, *ottobre* 19, *novembre* 23 en *dicembre* 13. Voor we hier nu voorbarige conclusies uit trekken over de sympatiekste maanden, moet gezegd worden dat deze frequentie aantoonbaar hoofdzakelijk uit het krantenmateriaal komt, en dat van de opgenomen kranten er twee uit mei, twee uit juni, één uit juli en één uit augustus stammen. Zo voorzichtig moet je met dit soort conclusies zijn. Hoe dan ook, niemand zal alleen de zomermaanden onderwijzen. Een ander punt is de spreiding van de woorden over de teksten; een goede frequentielijst betreft deze in zijn resultaten. Hoe dat hier zit, daar kom ik nog op terug. De beschikbaarheid (*disponibilità*) is een ander criterium, ontdekt door Michéa bij het werk aan het *Français fondamental*. Woorden als *mes*, *vork* en *lepel* zijn onmiddellijk beschikbaar voor alle taalgebruikers, maar dergelijke concrete zelfstandige naamwoorden komen in een frequentielijst praktisch niet voor, omdat er niet dagelijks over gepraat wordt. De cijfers van de IBM-lijst zijn *coltello* 18, *cucchiaio* 8, en *forchetta* helemaal niet, waarbij opgemerkt moet worden dat van de 18 gevallen van *coltello* er acht uit de kranten komen, hetgeen te denken geeft. Het bezwaar van het ontbreken van dit type woorden kan ondervangen worden door naast de frequentielijst met 'centres d'intérêt' te werken. Een volgend criterium is het gemak waarmee van een bepaald woord doorzichtige afleidingen en samenstellingen gemaakt kunnen worden. Een woord als *altoparlante* 'luidspreker', dat niet in de lijst voorkomt, kan best ingevoerd worden in een beginnerscursus, omdat het voor een Nederlander op geheel doorzichtige wijze gevormd is uit de zeer frequente woorden *alto* en *parlare*. Beperking van de lijst kan verkregen worden door synoniemen te verwijderen. Zo bevat de aangeboden lijst zowel *cominciare* als *incominciare*, zowel *sentire* als *udire*, zowel *succedere* als *accadere* en *avvenire*. Aangezien het meest frequente werkwoord steeds hetzelfde kan doen als de andere, is het voor de hand liggend de andere te schrappen, zeker voor het actief gebruik.

Ten slotte zijn sommige woorden domweg gemakkelijker te leren dan andere, en ook dat is een belangrijk criterium. Dat *zak sacco* is en *muur muro* hoef je de leerling maar een keer te vertellen, en dat medicijnen *medicina* heet en wiskunde *matematica* dat zal ook niemand verbazen. Omgekeerd stellen woorden als *scappare*, *strappare*, *scacciare*, *schiacciare*, *spaccare*, *spacciare* en *spicciarsi* de beginner voor grote problemen omdat ze voor hem evenveel op elkaar lijken als voor ons de namen van de Chinese leiders of van Japanse filmacteurs.

Om terug te keren tot de frequentie: een goede frequentielijst moet uitgaan van een homogeen corpus dat voldoende verscheiden is om het domineren van bepaalde thematische woorden, dat wil zeggen woorden die gebonden zijn aan een bepaald onderwerp, te vermijden. Bovendien moet het corpus, als de lijst een didactisch doel heeft, overeenstemmen met het soort taalgebruik dat men wil onderwijzen. Goede

voorbeelden zijn de lijst van het *Français fondamental* en de lijst van Šteinfeldt voor het Russisch. *Le français fondamental* gaat uit van 163 teksten gesproken taal, meest interviews en verder gesprekken afgeluisterd met een verborgen microfoon, en dergelijke. Deze 163 teksten bevatten 312.135 woorden, dat wil zeggen gemiddeld 1915 woorden per tekst (oftewel zeseneenhalf getikte vellen). Het aantal verschillende woorden dat zij bevatten is 7.995. De auteurs zijn echter van mening dat de rangorde om statistische redenen slechts betrouwbaar is tot een frequentie van maximaal 20, hetgeen hun 1063 woorden oplevert, waaronder ± 225 werkwoorden. Aangezien in de hier uitgedeelde lijst ook 300.000 woorden zijn verwerkt, is dit getal van 225 interessant. Bij ieder woord van *Le français fondamental* staat de spreiding vermeld, dat wil zeggen in hoeveel van de 163 teksten het voorkomt. De resultaten worden representatief geacht voor het gesproken Frans. De lijst van Šteinfeldt, bestemd voor Estlandse middelbare scholen, is gebaseerd op 400.000 woorden, gekozen uit 350 teksten, elk van 1000 tot 2000 woorden. Deze teksten komen uit vijf categorieën, namelijk kinderboeken, jeugdtijdschriften, toneelstukken, Russische uitzendingen van de Estlandse radio en krantenartikelen. Het totaal aantal verschillende woorden is 24.224. Hiervan zijn in de lijst alle woorden opgenomen die minstens eenmaal in 14 teksten voorkomen, dat zijn er 2500, waaronder naar schatting ± 550 werkwoorden. Het selectie criterium is hier dus eerder de spreiding dan de frequentie. Betrouwbaar is deze lijst, volgens de statistische berekeningen van Frumkina, tot en met 1100 tot 1300 woorden, dat wil zeggen tot een frequentie van 30; in feite betekent frequentie 30 vanwege de foutenmarge een frequentie tussen de 15 en de 45. Hierbij zitten bijna 300 werkwoorden. De resultaten lijken representatief voor het Sowjet-schoolkinderenproza. Gesteld dat onze lijst eenzelfde graad van betrouwbaarheid zou hebben als beide genoemde lijsten, dan betekent dit dat we ook hier een frequentie van 30 ongeveer als onderste drempel aan moeten houden en dat vanwege de foutenmarge hierbij nog een aantal uit de volgende honderd kunnen komen, maar waarschijnlijk praktisch geen werkwoorden met een frequentie onder de twintig. (Dit is een ruwe schatting).

De frequentielijst van de IBM is in zijn opzet een getrouwe navolging van de serie frequentiewoordenboeken van de Amerikaanse romanist Alphonse Juilland. Deze serie, waarvan een Spaans en een Roomeens deel zijn verschenen,¹ heeft een heel ander doel dan het taalonderwijs, zoals door Juilland ook uitdrukkelijk wordt gezegd. Het gaat hier om een vergelijking van de woordenschatopbouw, de fonologische structuur en dergelijke, tussen de Romaanse talen onderling en daarom is er allereerst gestreefd naar een nauwkeurige overeenkomst in de criteria die de keuze van het materiaal hebben bepaald. Juilland nam een scherp omliggende periode, namelijk het tijdperk tussen de twee wereldoorlogen, en koos daaruit zo objectief mogelijk teksten van vijf categorieën, te weten Toneel, Romans en novellen, Essayistiek,

¹ Het Italiaanse deel is van 1973, een jaar nadat deze recensie geschreven werd.

Technische literatuur, Kranten en tijdschriften. In iedere categorie werden door een expert ongeveer 50 teksten vastgesteld, waaruit volgens een vaste regel een aantal zinnen werd geselecteerd (bijvoorbeeld zin 1 van pagina 1, zin 2 van pagina 4, zin 3 van pagina 7, enzovoort). In totaal bevat iedere categorie 100.000 woorden. De lijst die het resultaat van deze werkwijze is gaat tot ongeveer de frequentie 5 en bevat 5000 woorden. Als we de eerder vermelde criteria toepassen, constateren we dat de manier waarop de zinnen gekozen werden een grote verscheidenheid garandeert binnen elk van de vijf categorieën (van elke roman werden bijvoorbeeld ± 80 zinnen gekozen, steeds drie pagina's van elkaar), dat echter de categorieën op zichzelf heterogeen zijn; de lijst is dus geen frequentielijst over een corpus van 500.000 woorden, maar een optelsom van vijf frequentielijsten over een corpus van 100.000 woorden per lijst. De omvang is ongeveer vergelijkbaar met die van de lijsten van *Le français fondamental* en Šteinfeldt, dat wil zeggen de frequentie is betrouwbaar tot zo'n 1500 woorden, *mits* het betrokken woord in alle categorieën gelijkelijk vertegenwoordigd is. Deze laatste voorwaarde is verdisconteerd in de spreidingscoëfficiënt: als een woord overal in gelijke frequentie voorkomt is de spreiding 100%; dit cijfer daalt bij verschillende frequentie volgens een formule die enerzijds gerelateerd is aan het verschil tussen de hoogste en de laagste frequentie per categorie, anderzijds aan de totale frequentie. Deze formule werkt, op voorwaarde dat alle categorieën gelijkwaardig zijn, dat wil zeggen dat een afwijkende frequentie in de rubriek Toneel even zwaar telt als een afwijkende frequentie in de rubriek Kranten. Als er, zoals we voor de IBM-lijst zullen zien, één categorie is die zich geregeld afwijkend gedraagt, dan werkt de formule niet meer. Tenslotte is, hoewel Juilland's doel niet didactisch was, het corpus representatief voor een passieve taalschat van de periode tussen de beide wereldoorlogen.

Tagliavini en zijn medewerkers hebben de opzet van Juilland geheel overgenomen, maar een aantal wijzigingen aangebracht.

1) Als periode is gekozen de tijd na de tweede wereldoorlog. Dit heeft tot nadeel dat de lijst niet meer vergelijkbaar is met die van Juilland; zo heeft bijvoorbeeld de Spaanse lijst een frequentie 2 voor *teléfono* in de romancategorie, de Italiaanse een frequentie 23 voor *telefono* in dezelfde categorie.² Maar aangezien de vergelijkbaarheid ook door de andere wijzigingen wordt aangetast, is de keuze van een recentere periode alleen maar een winstpunt.

2) De categorieën Toneel (T), Romans (R) en Kranten (G) zijn gehandhaafd; wel is de rubriek Giornali iets heterogener samengesteld, te weten 4 nummers van de Corriere della Sera, 1 nummer van Oggi en 1 van Quattrosoldi (dat wil zeggen het blad van de consumentenbond). De categorie Essayistiek is vervangen door de categorie Cinema (C) (dat wil zeggen filmscenario's), hetgeen beschouwd kan worden als een winstpunt, omdat dit het aandeel van de gesproken taal vergroot en een grotere aan-

² In de Italiaanse 'Juilland' ontbreekt het woord *telefono*; het adjectief *telefonico* komt er 10x voor, uitsluitend in de rubrieken Kranten en tijdschriften (4x) en Technische literatuur (6x).

sluiting met de omgangstaal mogelijk maakt dan de rubriek Teatro. Men vergelijk het woord *ciao*, totaalrequentie 161, deelfrequentie T 47, R 7, C 106, G 0, S 1. Niettemin maakt deze wijziging het corpus meer homogeen. De categorie Technische literatuur is vervangen door een categorie *Sussidiari* (S), dat wil zeggen leerboeken voor de lagere school die allerlei soort leerstof bevatten: religie, geschiedenis, aardrijkskunde, biologie, taal- en rekenlesjes. Deze keuze is ronduit ongelukkig; in de eerste plaats is het akelig materiaal, een zoetelijk, religieus en moralistisch getint proza, waarvan ik niet aanneem dat een Nederlandse leerling er met plezier naar zou grijpen; ik heb er een paar van bekeken en herinner me als afschrikwekkend voorbeeld de beschrijving van het leven van de bijen, die eindigt met een hooggestemde moraal over het voorbeeld van eendracht en samenwerking dat zij aan de mensen kunnen geven. In de tweede plaats, en dat is belangrijker: de teksten zijn zodanig gespecialiseerd van aard dat de woordenschat in sterke mate thematisch is, en dat zelfs in zeer hoge frequenties. Mijn favoriete voorbeeld is de frequentie van *sacerdote*. naast die van *prete*: Prete totaal 48, T 11, R 25, C 7, G 3, S 2. Sacerdote totaal 66, T 3, R 0, C 2, G 1, S 60. waarbij *prete* in de ranglijst van de afdalende frequentie ongeveer bij de 940 zit, *sacerdote* bij de 710, terwijl het toch kennelijk een woord is dat alleen in schoolboekjes wordt gebruikt. Met andere woorden, de frequenties van S zijn zo afwijkend dat ze het totaalbeeld vertroebelen.

3) Een derde wijziging, ook in ongunstige zin, is dat de aantallen teksten belangrijk kleiner zijn: in plaats van 50 romans zijn er 10 romans gebruikt, wat betekent dat het aantal gebruikte zinnen per roman vervijfvoudigd is, zodat het risico van thematische woorden groter is geworden. Het zal wel gedeeltelijk toeval zijn dat naast *romano*, *milanese*, *genovese* en *fiorentino* ook *ferrarese* vertegenwoordigd is, terwijl *torinese*, *bolognese* en *napoletano* ontbreken, maar ik kan niet nalaten een verband te leggen tussen het feit dat 2 van de 5 gevallen van *ferrarese* uit de romans stammen en de aanwezigheid van *Il Giardino dei Finzi-Contini*, dat in die stad speelt, in het corpus.

4) Een vierde wijziging, ook onplezierig, is een onvoldoende uitsorteren van de vormen. Terwijl Juilland voor het Spaans zorgvuldig onderscheidt tussen *cantaba* eerste persoon en *cantaba* derde persoon, is het bij de IBM-lijst niet mogelijk na te gaan of bijvoorbeeld *prese* passato remoto is of vrouwelijk meervoud van het participium. Voor de rangschikking in lemmata houdt de IBM zich aan het woordenboek van Migliorini-Cappuccini en dit heeft nare consequenties, speciaal voor het onderscheid tussen participia en adjectieven. Zo wordt het beeld van de werkwoordsfrequentie vervormd door de frequentie van het werkwoord *scorrere*, t.w. 106, waarvan er 70 op rekening van de vormen van *scorso* 'afgelopen, voorbij' komen (Ze komen haast allemaal in de krantenrubriek voor). Soortgelijke problemen rijzen met *morto*, *nato*, *innamorato* en dergelijke.

Resumerend kan ik stellen dat Tagliavini c.s. helaas het model van Juilland hebben gekozen (helaas, omdat hun lijst dus ook een conglomeraat van vijf categorieën is), dat zij helaas een aantal bedenkelijke wijzigingen hebben aange-

bracht, dat sommige wijzigingen evenwel toch groot nut hebben, speciaal vanwege het grotere accent op de spreektaal, en dat het geheel ondanks alle bedenkingen een belangrijke bijdrage tot de Italianistiek is.

Met het materiaal van de lijsten kun je allerlei aardige spelletjes doen. Ik heb bijvoorbeeld de categorieën uitgesplitst en voor elke categorie de ranglijst van de eerste 50 werkwoorden vastgesteld. Het resultaat is verbluffend. De werkwoorden van de eerste drie categorieën zijn praktisch dezelfde; de rangorde van de individuele werkwoorden toont geen substantiële verschillen en zo er in één rubriek al een werkwoord van de andere ontbreekt, dan komt dat toch onmiddellijk na de 50. Het voornaamste verschil is dat de rubriek Romans lagere absolute getalswaarden heeft (hetgeen neerkomt op een rijker vocabulaire) en verder dat bepaalde werkwoorden er een aantal plaatsen hoger staan, in de eerste 50 zijn dat in het bijzonder *rispondere* en *aprire*, in het algemeen zijn dat de werkwoorden van 'zeggen' en de werkwoorden die een beweging of gebaar beschrijven, en dat was om inhoudelijke redenen uiteraard te verwachten. De rubriek Sussidiari daarentegen geeft merkwaardige frequenties te zien; de meest opmerkelijke zijn: rangnummer 16 *formare*, 18 *sorgere*, 21 *osservare*, 22 *nascere* (plus *nato*), 24 *combattere*, 25 en 26 *abitare* en *costruire*, 30 *comprendere* (met *compreso*), 31 *giungere*, 37 *indicare*, 38 *costituire*, 39 *occupare*, 40 *lavorare*. Het meest welsprekend is het geval van *sorgere*: totaal frequentie 100, T 4, R 1, C 2, G 2, S 91, of dat van *osservare*: totaal frequentie 130, T 7, R 18, C 2, G 17, S 86. De rubriek Giornali heeft soortgelijke afwijkingen, alleen in mindere mate. Dit feit is opmerkelijk, want over het algemeen heerst de mening dat de thematische woorden te vinden zijn onder de zelfstandige naamwoorden, maar dat de belangrijkste werkwoorden een vrij stabiele frequentie hebben.

Ik heb er de volgende consequentie uit getrokken. Het heeft geen zin te werken met de totale frequentie van de IBM-lijst, aangezien dit een cumulatie van heterogeen materiaal zou zijn. Daarentegen heeft het wel zin om de frequenties van Teatro, Romanzi en Cinema te combineren, aangezien die gegevens elkaar versterken. Het resultaat is een lijst waarin de gesproken taal domineert en dat zou juist een lijst zijn die nut heeft voor een actief te beheersen beginnerswoordenschat. Het vocabulaire van de romans verschilt, naar mij bij het opstellen van de lijst gebleken is, niet al te zeer van dat van Teatro en Cinema; de verschillen treden vooral op bij de relatieve frequentie van de personen en tijden van de werkwoorden, en verder zijn er bij de lagere frequenties een aantal typische romanwoorden; een duidelijk geval is het werkwoord *apparire*, waarvan de cijfers zijn: totaal 107, T 6, R 37, C 3, G 34, S 27 (totaal TRC 46, rangnummer 150)

Even een paar getallen: het totale aantal werkwoorden in de Piccolo Garzanti is 5249. In de lijst van de IBM zijn ruim 1250 werkwoorden opgenomen; hiervan hebben er 324 een frequentie van 20 of meer in de categorieën Teatro, Romanzi en Cinema, en ongeveer 550 een frequentie van 10 of meer. Zuiver intuïtief acht ik de frequenties beneden de 10 volkomen willekeurig.

Het heeft me aan de tijd ontbroken om datgene te doen wat ik oorspronkelijk van plan was, namelijk definitieve conclusies te trekken ten aanzien van de allereerst te onderwijzen werkwoordsvormen. Ik wil wel enkele voorlopige conclusies geven:

1) Van de werkwoorden op *-ire* hebben de werkwoorden zonder *-isco* over het algemeen een hoge frequentie; van die met *-isco* zijn de belangrijkste *capire* en *finire*, waarbij de voorsprong van *capire* zodanig is dat het te overwegen zou zijn om dat werkwoord als paradigmavoorbeeld te geven (het enige bezwaar is de geringe frequentie van de imperatief).

2) De groep *avere-sapere-fare-andare-stare-dare* is zo frequent dat het overweging verdient de *presente*-vormen hiervan dadelijk aan te bieden (dit geldt ook voor *dovere*, *potere*, *volere* en *essere*).

Van de overige onregelmatige *presenti* zijn de volgende belangrijk:

a) Van de infiniti contratti: *fare*, *dire*, *bere*.

b) Van de *-go*-groep: *venire*, *tenere*, *rimanere*, *salire*, *togliere scegliere*; *rimanere* kan hierbij als paradigmavorm gelden.

c) Het werkwoord *parere* is alleen belangrijk als onpersoonlijk werkwoord (op een totaal van 242, heeft *pare* een frequentie 140, *pareva* 71 en het geheel van de derde-persoonsvormen 233).

d) Van *piacere* is de vorm *piacciono* belangrijk; *tacere* en *giacere* horen niet tot de frequente werkwoorden.

e) De werkwoorden *uscire* en *riuscire* zijn belangrijk.

3) Verreweg de meest frequente onregelmatige *futuro*-vormen zijn de volgende negen: *sarò* 994, *avrò* 562, *potrò* 320, *dovrò* 272, *vorrrò* 252, *vedrò* 111, *verrò* 67, *andrò* 66, *saprò* 48. Daarna komt *rimarrò* met 12. Deze cijfers hebben betrekking op alle *futuro*- en *condizionale*-vormen voor alle vijf categorieën en laten duidelijk het verband zien tussen frequentie en onregelmatige vorm.

Rest mij nog de gegevens te vermelden over de aan het begin genoemde werkwoorden. Deze zijn:

<i>ballare</i>	rang 176	frequentie 38
<i>baciare</i>	rang 179	frequentie 37
<i>badare</i>	rang 190	frequentie 35
<i>bagnare</i>	2e groep	frequentie 15
<i>balbettare</i>		frequentie 8

Appendix

De oorspronkelijke tekst bevatte als appendix een lijst van alle werkwoorden met een frequentie groter dan 20. Hier worden alleen de eerste vijftig werkwoorden opgenomen.

1	essere	12.987	26	aspettare	347
2	avere	6.371	27	portare	332
3	fare	3.606	28	conoscere	297
4	dire	2.538	29	sembrare	268
5	andare	1.899	30	finire	265
6	volere	1.833	31	ricordare	261
7	sapere	1.704	32	arrivare	243
8	potere	1.406	33	parere	242
9	vedere	1.312	34	piacere	242
10	stare	1.243	35	chiamare	241
11	dovere	1.212	36	tenere	238
12	venire	1.015	37	cercare	236
13	dare	777	38	perdere	224
14	sentire	612	39	chiedere	223
15	capire	607	40	entrare	204
16	parlare	565	41	riuscire	203
17	mettere	533	42	bisognare	192
18	prendere	510	43	bastare	180
19	credere	496	44	rispondere	176
20	guardare	494	45	restare	176
21	lasciare	494	46	sposare	167
22	pensare	474	47	mangiare	166
23	trovare	408	48	uscire	162
24	tornare	388	49	vivere	161
25	passare	354	50	scrivere	160

Bibliografie

- Bortolini, Umberta, Carlo Tagliavini & Antonio Zampolli. *Lessico di frequenza della lingua italiana contemporanea*. Milano: IBM Italia, 1971.
- Brunet, Gilbert et René La Borderie. *Insieme a Venezia*. Paris: A. Colin, 1966.
- Gougenheim, Georges. *Dictionnaire fondamental de la langue française*. Paris: Didier, 1958.
- _____, René Michéa, Paul Rivenc et Aurélien Sauvageot. *L'élaboration du français fondamental*. Paris: Didier, 1956.
- Il piccolo dizionario Garzanti della lingua italiana*. Milano: Garzanti.
- Juilland, Alphonse. *Frequency Dictionary of Spanish Words*. The Hague: Mouton, 1964.
- _____, & Vincenzo Traversa, *Frequency Dictionary of Italian Words*. The Hague: Mouton, 1973.
- Mackey, William F. *Language Teaching Analysis*. London: Longmans, 1969.
- Matoré, Georges. *Dictionnaire du vocabulaire essentiel* (les 5000 mots fondamentaux). Paris: Larousse, 1963.
- Pestelli, Leo. *Parlare italiano*. Milano: Longanesi, 1962.
- Šteinfeldt, Evi. *Russian Word Count. 2500 Words Most Commonly Used in Modern Literary Russian*. Moscow: Progress Publishers, z.d. (= ±1962).