

L'ANALISI LINGUISTICA E LETTERARIA

FACOLTÀ DI SCIENZE LINGUISTICHE E LETTERATURE STRANIERE
UNIVERSITÀ CATTOLICA DEL SACRO CUORE

1

ANNO XXV 2017

MARE PUNICVM.

MARE LIBICVM

EDUCATT - UNIVERSITÀ CATTOLICA DEL SACRO CUORE

L'ANALISI
LINGUISTICA E LETTERARIA

FACOLTÀ DI SCIENZE LINGUISTICHE
E LETTERATURE STRANIERE

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

1

ANNO XXV 2017

PUBBLICAZIONE SEMESTRALE

L'ANALISI LINGUISTICA E LETTERARIA
Facoltà di Scienze Linguistiche e Letterature straniere
Università Cattolica del Sacro Cuore
Anno XXV - 1/2017
ISSN 1122-1917
ISBN 978-88-9335-209-3

Direzione

LUISA CAMAIORA
GIOVANNI GOBBER
LUCIA MOR
MARISA VERNA

Comitato scientifico

ANNA BONOLA – LUISA CAMAIORA – ARTURO CATTANEO – SARA CIGADA
ENRICA GALAZZI – MARIA CRISTINA GATTI – MARIA TERESA GIRARDI
GIOVANNI GOBBER – DANTE LIANO – MARIA LUISA MAGGIONI
GUIDO MILANESE – FEDERICA MISSAGLIA – LUCIA MOR – AMANDA MURPHY
FRANCESCO ROGNONI – MARGHERITA ULRYCH – MARISA VERNA
SERENA VITALE – MARIA TERESA ZANOLA

Segreteria di redazione

SARAH BIGI – ELISA BOLCHI
ALESSANDRO GAMBA – GIULIA GRATA

*I contributi di questa pubblicazione sono stati sottoposti
alla valutazione di due Peer Reviewers in forma rigorosamente anonima*

© 2017 EDUCatt - Ente per il Diritto allo Studio universitario dell'Università Cattolica
Largo Gemelli 1, 20123 Milano | tel. 02.7234.2235 | fax 02.80.53.215
e-mail: editoriale.dsu@educatt.it (*produzione*); librario.dsu@educatt.it (*distribuzione*)
web: www.educatt.it/libri

Redazione della Rivista: redazione.all@unicatt.it | web: www.analisilinguisticaeletteraria.eu

Questo volume è stato stampato nel mese di luglio 2017
presso la Litografia Solari - Peschiera Borromeo (Milano)

INDICE

Articulations of the Economic Motif in Shakespeare's <i>Romeo and Juliet</i> <i>Luisa Camaiora</i>	7
Charity, Melancholy, and the Protestant Ethic in Herman Melville's <i>Bartleby</i> and <i>Cock-A-Doodle-Do!</i> <i>Federico Bellini</i>	29
La lingua di Internet in Russia: stato della ricerca <i>Laila Paracchini</i>	45
Come fare le cose con i testi: <i>A Modell of Christian Charity</i> di John Winthrop <i>Carla Vergaro</i>	99
'Écologie' et 'environnement' dans l'espace dictionnaire français <i>Michela Murano</i>	117
Forme di espressione della causalità nel confronto francese-tedesco <i>Sibilla Cantarini e Gaston Gross</i>	131
Estrazione del 'che' polivalente da un corpus POS-taggiato: limiti e possibilità <i>Marco Budassi</i>	147
Definitional Arguments in Children's Speech <i>Rebecca Schär</i>	173
Tra stabilità sociale e pornografia: giochi di parole sovversivi e armonizzazione su Internet nell'era Hu Jintao <i>Nazarena Fazzari</i>	193
Recensioni e Rassegne	
Recensioni	217
Rassegna di Linguistica generale e di Glottodidattica a cura di Giovanni Gobber	231

Rassegna di Linguistica francese a cura di Enrica Galazzi e Chiara Molinari	241
Rassegna di Linguistica inglese a cura di Amanda Murphy e Margherita Ulrych	249
Rassegna di Linguistica russa a cura di Anna Bonola	257
Rassegna di Linguistica tedesca a cura di Federica Missaglia	261
Rassegna di Tradizione della cultura classica a cura di Guido Milanese	267
Indice degli Autori	273

NOTA INTRODUTTIVA

Siamo lieti di inaugurare in questo fascicolo la nuova Rassegna di Tradizione della cultura classica, dedicata alla segnalazione di opere recenti relative al rapporto tra la cultura classica e tardoantica e la cultura moderna e contemporanea. Le schede saranno redatte preferibilmente in inglese, ma saranno accolti i contributi nelle più diffuse lingue europee. Ci si augura che questa Rassegna possa costituire un momento di contatto tra studiosi che, movendo da specializzazioni diverse, si riconoscono nella tradizione della cultura europea.

I Direttori con Guido Milanese

ESTRAZIONE DEL 'CHE' POLIVALENTE DA UN CORPUS POS-TAGGATO: LIMITI E POSSIBILITÀ

MARCO BUDASSI

Cifra caratteristica dell'italiano neo-standard sono tratti linguistici una volta appartenenti al sub-standard, come il 'che' polivalente. In questo lavoro mi sono occupato di valutare se sia possibile o meno rintracciare il 'che' polivalente in un corpus annotato solamente per *part of speech*. Dai dati emerge come le sotto-tipologie di 'che' polivalente di tipo 'giustappositivo' siano estraibili dal corpus tramite *query*, mentre le sotto-tipologie di tipo 'restrittivo' non lo siano.

To the so-called neo-standard Italian language belong linguistic features that used to be considered as sub-standard. One of these features is the polyvalent use of 'che'. In this paper, I aimed at evaluating whether this polyvalent use of 'che' may be traced within a part of speech-tagged corpus or not. Data show that 'appositive' sub-types of 'che' can be mined from the corpus via queries, while 'restrictive' sub-types cannot.

Keywords: corpus linguistics, POS-tagging, polyvalent use of *che*, neo-standard Italian, syntax

1. Introduzione: l'italiano contemporaneo e il 'che' polivalente

In prospettiva sociolinguistica sono state proposte innumerevoli modellizzazioni dell'articolazione in varietà dell'italiano contemporaneo¹. Per fare un esempio, Berruto² ha suddiviso l'italiano contemporaneo in ben nove varietà³. Queste sono individuate sulla base di alcuni parametri definitori ormai ampiamente accettati in letteratura⁴, ovvero diatopia, diastratia, diafasia e diamesia.

In realtà, la sociolinguistica dell'italiano contemporaneo è in continua evoluzione⁵. Da circa vent'anni a questa parte sono in atto alcune tendenze innovative della lingua, come

¹ Si vedano, ad esempio: G.B. Pellegrini, *Saggi di linguistica italiana. Storia, struttura, società*, Boringhieri, Torino 1975; A. Mioni, *La situazione sociolinguistica italiana: problemi di classificazione e di educazione linguistica*, in *L'educazione linguistica. Atti della giornata di studio GISCEL*, P. Benincà et al. ed., Cleup, Padova 1975, pp. 29-34; Id., *La situazione sociolinguistica italiana: lingua, dialetti, italiani regionali*, in *Guida all'educazione linguistica*, A. Colombo ed., Zanichelli, Bologna 1979, pp. 101-114; T. De Mauro, *Profilo linguistico della società italiana*, in *Lingua e dialetti nella cultura italiana da Dante a Gramsci*, S. Gensini – T. De Mauro – T. Leto ed., G. D'Anna, Messina/Firenze 1980.

² G. Berruto, *Sociolinguistica dell'italiano contemporaneo*, Carocci Editore, Roma 2012.

³ Italiano (it.) standard letterario, it. neo-standard, it. parlato colloquiale, it. regionale popolare, it. informale trascurato, it. gergale, it. formale aulico, it. tecnico-scientifico e it. burocratico.

⁴ Si veda, ad esempio, E. Coseriu, *Lezioni di linguistica generale*, Boringhieri, Torino 1973.

⁵ A.A. Sobrero – A. Miglietta, *Lingua e dialetto nell'Italia del Duemila*, Congedo, Galatina 2006.

l'aumento della variazione diafasica visto il nascere di una nuova classe di situazioni d'uso create dalla comunicazione digitale. Oltre a ciò, l'italiano contemporaneo mostra una netta diminuzione della differenziazione diastratica e diatopica. Questo è spiegabile anche sulla base della 'ristandardizzazione' dell'italiano in corso al giorno d'oggi, che tende a promuovere tratti linguistici un tempo diastraticamente bassi. A questo orientamento probabilmente contribuisce anche il fatto che in situazioni colloquiali oggi si tenda più a cambiare varietà linguistica che lingua, mentre fino a poco tempo fa si attuava *code switching* al dialetto.

Queste tendenze hanno portato, secondo alcuni autori⁶, a un allargamento sostanziale dell'italiano standard a tratti che fino a pochi decenni fa appartenevano al sub-standard e venivano identificati dalle grammatiche normative come errati. Tale processo è quindi giunto a identificare una nuova varietà di italiano, la quale è stata definita 'neo-standard'. Certo, la definizione di cosa sia lingua 'standard' o 'neo-standard' è di per sé complessa da affrontare. Secondo Berruto⁷, l'italiano neo-standard ha caratteristiche molto vicine a quelle dell'italiano standard, ma è caratterizzato da una maggiore tendenza verso il parlato colloquiale (a livello diamesico)⁸, verso l'italiano popolare (a livello diastratico) e verso l'italiano informale o settoriale (a livello diafasico). Questa differenziazione può essere chiarita con alcuni esempi. Sotto l'etichetta di italiano standard rientra l'italiano colto della tradizione letteraria, formale, e sovraregionale. Il suo uso è ormai ristretto solo a determinate classi professionali (come attori, annunciatori, doppiatori) o allo scritto formale⁹. Con italiano neo-standard, invece, s'intende l'italiano scritto impiegato nella prosa giornalistica, in buona parte di quella letteraria, nello scritto informale e nel parlato informale. È dunque interessante notare, seguendo Alfonzetti¹⁰, come la ristandardizzazione dell'italiano contemporaneo non abbia in realtà intaccato le radici profonde del sistema, ma abbia determinato un diverso rapporto di forze tra uso e norma.

L'esistenza di un italiano neo-standard contrapposto all'italiano standard non è però unanimemente accettata dagli studiosi¹¹. A tal proposito, secondo Berruto¹², una risposta più precisa in merito a tale questione potrà forse essere raggiunta attraverso lo studio di dati empirici e analisi di *corpora*. Una tale operazione potrebbe infatti aiutare a comprendere diffusione e comportamento di gran parte di quei fenomeni linguistici che sono tipici dell'italiano neo-standard. Simili studi, tuttavia, non sono ancora numerosissimi¹³.

Proprio in riflessioni di tal genere ha origine la motivazione di questo lavoro. Nel complesso, i mutamenti in atto e caratterizzanti l'italiano neo-standard sono numerosi e coin-

⁶ Es. G. Berruto, *Sociolinguistica dell'italiano contemporaneo*.

⁷ *Ibidem*.

⁸ Si veda anche P. V. Mengaldo, *Storia della lingua italiana. Il Novecento*, Il Mulino, Bologna 1994.

⁹ C. Andorno – P. Ribotta, *Insegnare e imparare la grammatica*, Paravia Scriptorium, Torino 1999.

¹⁰ G. Alfonzetti, *La relativa non-standard – italiano popolare o italiano parlato?*, Centro di studi filologici e linguistici siciliani, Palermo 2002.

¹¹ Si veda, ad esempio, L. Serianni, *Il problema della norma linguistica dell'italiano*, "Annali dell'Università per Stranieri di Perugia", 7, 1986, pp. 47-69.

¹² G. Berruto, *Sociolinguistica dell'italiano contemporaneo*.

¹³ G. Alfonzetti, *La relativa non-standard – italiano popolare o italiano parlato?*

volgono tutti i livelli linguistici¹⁴. Tra di essi, a livello di sintassi, compare il 'che' polivalente. Volendo seguire la linea tracciata da Berruto¹⁵, in questo lavoro mi occuperò di valutare se sia possibile o meno rintracciare in un corpus annotato per *part of speech* (POS)¹⁶ il 'che' polivalente. Per porre la questione in termini più specifici, scopo di questo contributo è capire se e come sia possibile individuare la percentuale di probabilità di polivalenza di un 'che' appartenente a un dato *pattern* POS-taggiato, estratto da un corpus tramite *query*. Una simile analisi, per quanto preliminare, appare necessaria nell'ottica di indagare fenomeni linguistici come il 'che' polivalente in prospettiva quantitativa.

Questo contributo è organizzato come segue. In primo luogo, cercherò di analizzare a livello qualitativo il 'che' polivalente; darò poi una definizione operativa del 'che' polivalente e fornirò una rassegna ordinata delle varie sotto-tipologie (Sezione 2). Dopo una sezione di metodologia (Sezione 3), passerò allo studio su un corpus di ogni tipo di 'che' polivalente descritto nella Sezione 2 (Sezione 4). Discuterò poi i risultati evidenziati dalla ricerca nel corpus, traendo alcune conclusioni su quante e quali tipologie di *che* polivalente possano essere individuate su base contestuale in un corpus taggato per *part of speech* (Sezione 5).

2. Analisi qualitativa

Definire cosa sia un 'che' polivalente è complesso, e ciò è dovuto al fatto che sotto l'etichetta di 'che polivalente' si trovano in realtà a coesistere un determinato numero di fenomeni linguistici molto differenti fra loro. Grossolanamente, si potrebbe bipartire l'insieme di questi fenomeni in un primo gruppo al quale appartengono casi di 'che' complementatore impiegato come generico introduttore di frasi (e.g. Che poi magari ci vanno lo stesso!) e un secondo gruppo a cui appartengono casi di che introduttore di frasi relative restrittive, le quali modificano un sintagma nominale (e.g. Maledetto il giorno che ti ho incontrato.). Dovendo però stabilire una base per condurre lo studio sul corpus in maniera efficace, mi vedo costretto a proporre una definizione operativa (senza alcuna pretesa di esaustività) del 'che' polivalente e una sua articolazione in tipologie che trascendano questa bipartizione. Dunque, in questo lavoro considererò polivalente ogni 'che' sostitutivo di altro. Questa definizione amplia il punto di vista di Fiorentino¹⁷, secondo la quale il 'che' diviene polivalente quando è utilizzato per introdurre frasi subordinate le quali, normalmente, in italiano sarebbero introdotte da altre congiunzioni subordinanti più precise.

¹⁴ Per una rassegna, si veda C. Andorno – P. Ribotta, *Insegnare e imparare la grammatica*, pp. 151-152.

¹⁵ G. Berruto, *Sociolinguistica dell'italiano contemporaneo*.

¹⁶ Ho scelto di focalizzarmi su un corpus taggato unicamente per *part of speech* (e non, per esempio, con funzioni sintattiche) nell'ottica di preparare il terreno per uno studio quantitativo del fenomeno. Infatti, il POS-tagging è di certo la tipologia di annotazione più rapidamente estensibile a grandi quantità di documenti testuali. Inoltre, in riferimento alle risorse disponibili ad oggi per la lingua italiana, la maggior parte di esse presenta solo un'annotazione per *part of speech* (e.g. CORIS corpus, Perugia corpus, La Repubblica corpus).

¹⁷ [http://www.treccani.it/enciclopedia/che-polivalente_\(Enciclopedia_dell'Italiano\)/](http://www.treccani.it/enciclopedia/che-polivalente_(Enciclopedia_dell'Italiano)/) ultima consultazione 27 agosto 2016.

Alcune criticità sono sollevate da una tale definizione. Ad esempio, il suo grado di generalizzazione non rende in realtà conto della differenza esistente tra i vari tipi di ‘che’ polivalente (o anche solo le due macro-categorie di ‘che’ generico introduttore di frase da un lato, *che* introduttore di frasi relative restrittive dall’altro) impiegati dalla lingua italiana. In secondo luogo, va rilevato come mentre alcuni casi di polivalenza del ‘che’ sono soggetti a valutazioni normative oscillanti (per esempio, una frase come ‘Vieni che ti pettino’ è talvolta considerata corretta a livello di italiano standard), altri sono con più decisione identificabili come ‘errati’ (e.g. ‘Il paese che sono stato domenica scorsa si chiama P.’)¹⁸. Questo ha portato Galli de’ Paratesi a collocare i vari usi del ‘che’ polivalente lungo un *continuum* di accettabilità¹⁹. Nonostante queste mancanze, tuttavia, la definizione operativa di ‘che’ polivalente sopra formulata appare funzionale ai fini di questo lavoro. Tale definizione, infatti, permette di identificare in maniera chiara un basilare tratto comune ai vari tipi di ‘che’ polivalente, ovvero l’intrinseco uso non-standard.

Il ‘che’ polivalente è un fenomeno (sintattico) sfuggente sul piano linguistico. Secondo Berruto²⁰, il dominio della sintassi può essere tripartito in sintassi del periodo, fenomeni riguardanti l’ordine dei costituenti frasali e fenomeni inerenti alla coesione sintattica delle frasi. Il ‘che’ polivalente rientra all’interno della prima sottocategoria. In questa prospettiva, il ‘che’ può avere valore esplicativo (1), può marcare una frase relativa (2) o può comportarsi da generico introduttore di frase (3).

- (1) [...] i vecchi romani lo chiamano la lasagna... che lasagna significa pasta al forno [...]²¹.
- (2) [...] medicina è una cosa che si studia un macello.
- (3) [...] in un paesino... della Sicilia, della Calabria... che... ancora vanno in giro coll’asinello [...].

Il ‘che’ polivalente è poi utilizzato come generico indicatore di subordinazione, come mostrano gli esempi (4)²², (5) e (6) tratti da Foresti, Morisi e Resca²³. Questo avviene generalmente nell’italiano popolare²⁴.

- (4) Si sentiva che era piovuto che era poco, lì.
- (5) [...] cominciamo a correre sopra la ferrovia fino a Poggiorusco, no, che c’erano 17 o 18 chilometri.

¹⁸ G. Alfonzetti, *La relativa non-standard – italiano popolare o italiano parlato?*

¹⁹ N. Galli de’ Paratesi, *Mutamenti sociali e norma linguistica*, in *L’italiano in America latina*, V. Lo Cascio ed., F. Le Monnier, Firenze 1987.

²⁰ G. Berruto, *Varietà dialesiche, diastratiche, diafasiche*, in *Introduzione all’italiano contemporaneo. La variazione e gli usi*, A.A. Sobrero ed., Laterza, Bari 1993, pp. 3-36.

²¹ Si parla di un coltello.

²² Il ‘che’ polivalente in questo caso è il secondo.

²³ F. Foresti – P. Morisi – M. Resca ed., *Era come a mietere: testimonianze orali e scritte di soldati sulla grande guerra con immagini inedite*, Strada maestra, San Giovanni in Persiceto 1982.

²⁴ G. Berruto, *Varietà dialesiche, diastratiche, diafasiche*.

- (6) Se mi dà una licenza che vado a Bologna.

In questi casi, il 'che' assume un valore subordinante molto debole. Questo significa che il 'che' non esplicita una chiara relazione sintattica tra la frase principale e la frase dipendente; al contrario, la frase dipendente introdotta da questo tipo di 'che' si limita a costituire uno sviluppo di tipo generico della frase principale. Sono poi attestati alcuni esempi del 'che' come introduttore di una frase indicante evento, senza più alcun valore di collegamento interfrasale (7)²⁵.

- (7) Che... dice che c'era lavoro.

Nell'italiano popolare, il 'che' è anche usato come rafforzativo di altre congiunzioni subordinanti, dando luogo a costruzioni come 'siccome che', 'mentre che', 'come che', 'malgrado che', ecc. Come evidenziato da Mauri e Giacalone Ramat²⁶, però, quest'uso del 'che' era già diffuso nell'italiano antico, seppure spesso legato a specifiche sfumature semantico-sintattiche. Chiari esempi sono dati da 'che' in posizione immediatamente successiva a 'però' quando 'però' introduce una subordinata causale ("Tuttavia, però che molte volte lo numero del nove ha preso luogo tra le parole dinanzi [...]", Dante Alighieri, *Vita nova*, c. 1292-93, cap. 28, 1-3), o da che in posizione immediatamente successiva a mentre quando mentre introduce una subordinata temporale ("E mentre che egli [...] men cautamente con le' scherzava, avvenne che l'abate [...] sentio lo schiamazzio", Giovanni Boccaccio, *Decamerone*, 1370, cap. 1, 4). Vista tale distribuzione, in questo lavoro non considererò il 'che' rafforzativo di altre congiunzioni subordinanti come un caso di 'che' polivalente.

Da ultimo, altro costrutto tipico dell'italiano popolare è quello in cui il 'che' relativizza qualunque posizione sintattica, come evidente da (8), (9) e (10), tratti da Foresti, Morisi e Resca²⁷.

- (8) [...] un forte chiamato il Calvario, che lo battezzò così Napoleone.
 (9) [...] dell'acqua che c'erano dei vermi.
 (10) [...] dentro al ricovero che mi aveva portato la signorina.

Da un punto di vista sociolinguistico, il 'che' polivalente è estremamente vario e coinvolge più dimensioni di variazione²⁸, eccezion fatta per quella diatopica²⁹. Tuttavia, secondo Sabatini³⁰, più che in relazione a vari parametri sociolinguistici il 'che' polivalente andrebbe

²⁵ C. Grassi – M. Pautasso, *Biellesi nel mondo. Prima roba il parlare. Lingue e dialetti dell'emigrazione biellese*, Electa, Milano 1989. Il *che* polivalente in questo caso è il primo.

²⁶ C. Mauri – A. Giacalone Ramat, *The development of adversative connectives: stages and factors at play*, "Linguistics", 50, 2012, 2, pp. 191-239.

²⁷ F. Foresti – P. Morisi – M. Resca ed., *Era come a mietere*.

²⁸ G. Alfonzetti, *La relativa non-standard – italiano popolare o italiano parlato?*

²⁹ R. Sornicola, *Sul parlato*, Il Mulino, Bologna 1981.

³⁰ F. Sabatini, *L'italiano dell'uso medio – una realtà tra le varietà linguistiche italiane*, in *Gesprochenes Italienisch in Geschichte und Gegenwart*, G. Holtus – E. Radtke ed., Gunter Narr Verlag, Tübingen 1985.

analizzato tenendo in considerazione la divergenza che è possibile rilevare tra norma e uso dell'italiano, o per meglio dire il conflitto esistente tra la norma stessa e i meccanismi di semplificazione linguistica tipici del secondo.

Tutto ciò può essere rapidamente verificato facendo riferimento alla *Grande Grammatica Italiana di Consultazione* a cura di Renzi, Salvi e Cardinaletti (2001). In questa grammatica, il 'che' polivalente non è trattato in maniera specifica. Essendo la *Grande Grammatica Italiana di Consultazione* una grammatica dell'uso, i curatori si limitano ad annoverare il 'che' tra gli introduttori di proposizioni relative, complete, ma anche temporali o causali. Sicuramente, il 'che' è nell'italiano d'oggi introduttore di un gran numero di tipi di frasi diversi. Ad ogni modo, per ricondurci alla stessa prospettiva della definizione operativa del 'che' polivalente proposta in precedenza, in molti casi questo avviene solamente perché 'che' si trova a sostituire un'altra congiunzione, locuzione, ecc. Se anche non si può parlare di 'correttezza' o 'non correttezza' grammaticale, quindi, visti i fini di questo lavoro mi sembra necessario riuscire a separare 'che' polivalenti e non, caso per caso, passando in rassegna i vari tipi di frase introdotte da 'che' e individuando in quali casi 'che' si trovi in realtà in sostituzione di un altro elemento linguistico.

Secondo Renzi, Salvi e Cardinaletti³¹, in italiano 'che' introduce proposizioni complete non infinitive (Voglio che ognuno faccia il proprio lavoro), delle quali fanno parte tanto le soggettive (Conviene che tu torni a casa adesso), quanto le oggettive (Abbiamo spiegato che siamo stanchi). Il che introduce inoltre interrogative dirette o indirette (Che/che cosa vuoi?/Il dubbio su che cosa gli fosse permesso lo paralizzava). Che introduce anche concessive (Che ti piaccia o no io parto), consecutive (Sono talmente stanco che non mi reggo più in piedi), comparative (Ho parlato meno con ragazzi che con ragazze) e dichiarative (Devi dire questo, che io gli ho scritto tutto).

Trattato a parte può essere il caso delle proposizioni temporali. Renzi, Salvi e Cardinaletti³² annoverano il 'che' tra gli introduttori anche di questo tipo di frase. In alcuni casi il 'che' non sembra polivalente, come quando è preceduto da 'prima' (Prima che Giovanni tornasse a casa Maria era già uscita); in altri, tuttavia, è meno chiaro. La problematicità di questo tipo di costruzioni è stata già sottolineata da Fiorentino³³. Nel caso di costituenti temporali obliqui, infatti, l'oscillazione anche nell'italiano standard tra il ricorso a un pronome relativo obliquo o al complementatore 'che' è evidente³⁴. Tuttavia, la prospettiva che adotterò in simili casi è differente da quella di Fiorentino. Un esempio utile da considerare è dato dalla locuzione 'tutte le volte che'. Secondo i criteri definitivi delineati in precedenza, questo caso potrebbe essere fatto ricadere sotto la polivalenza del 'che' ('che' potrebbe sostituire il pronome relativo obliquo 'in cui'). In realtà, non considererò in questo lavoro

³¹ L. Renzi – G. Salvi – A. Cardinaletti ed., *Grande Grammatica Italiana di Consultazione*, Il Mulino, Bologna 2001, da cui sono tratti gli esempi che seguono.

³² *Ibidem*.

³³ G. Fiorentino, *Relativa debole*, Franco Angeli, Milano 1999.

³⁴ Secondo Fiorentino, questo può essere dovuto all'uso standard di 'che' nella relativizzazione di un complemento di tempo continuato (e.g. Le cinque ore che dura lo spettacolo), la quale renderebbe più accettabile la relativizzazione con 'che' anche di complementi di tempo di altro tipo (e.g. L'ultimo giorno che si può presentare la richiesta è il 31 maggio).

un simile uso del 'che' come polivalente. 'Tutte le volte che' è una locuzione temporale ormai cristallizzata; il 'che' non ha quindi valore sostitutivo rispetto a nulla, se anche il confine rimane sottile. Un altro esempio efficace è portato dal confronto di un caso come 'Sono cinque giorni che non ho più notizie di Massimo' con un caso come 'Maledetto il giorno che ti ho incontrato'. Seppure le due frasi siano (anche sintatticamente) abbastanza simili, considererò solamente la seconda di esse come contesto in cui occorre un 'che' polivalente, in quanto 'che' sostituisce 'in cui' abbastanza chiaramente. A partire da questo esempio è intuibile come la definizione di 'che' polivalente alla base di quest'analisi non sia fondata su criteri sociolinguistici, ma semantico-sintattici. 'Maledetto il giorno che ti ho incontrato' mostra infatti un uso di 'che' ampiamente accettato in italiano neo-standard, al pari di 'Sono cinque giorni che non ho più notizie di Massimo'. Tuttavia, nel primo caso 'che' è sostitutivo di 'in cui', nel secondo no. Conseguentemente, il primo 'che' verrà considerato polivalente, il secondo no.

Renzi, Salvi e Cardinaletti³⁵ considerano il 'che' introduttore anche di frasi di tipo causale o di protasi in costruzioni ipotetiche. Nel primo caso (È venuto con il treno, che aveva la macchina da aggiustare), però, esso si trova in sostituzione di 'poiché/perché', mentre nel secondo è operatore condizionale in sostituzione di 'se'. Secondo i criteri definitori che ho stabilito per questo lavoro, dunque, 'che' come questi o simili verranno considerati polivalenti.

Un'altra tipologia di 'che' sulla quale è necessaria una riflessione specifica è quella del 'che' definibile come 'enfaticante' o 'esclamativo' (Che sogno che ho fatto)³⁶. Questo tipo di 'che' potrebbe essere considerato polivalente perché sembra indicare solo uno sviluppo generico, con valore subordinante molto debole (esattamente come nei casi 4-6). Tuttavia, da un lato questo tipo di 'che' mal si colloca entro la definizione di 'che' polivalente che è stata data in precedenza, dall'altro può semplicemente essere considerato un introduttore di frase esclamativa.

La problematicità principale legata a questo tipo di 'che' sta nel fatto che è difficile specificare in sostituzione di cosa esso potrebbe trovarsi. Una buona ipotesi potrebbe essere: in contesti come 'Che sogno che ho fatto!', 'che' sostituisce 'quale' (Quale sogno...). Tuttavia, in casi simili 'che' e 'quale' non sono perfettamente interscambiabili. Si prenda ad esempio il contesto 'Che strano che è tutto ciò!'. Secondo Renzi, Salvi e Cardinaletti³⁷, il primo 'che' è esclamativo, mentre il secondo è un semplice complementatore. Il secondo 'che' (il complementatore), però, è in distribuzione complementare rispetto a 'quale': esso può infatti essere inserito solo in dipendenza da un 'che' introduttore di frase (Che sforzi che ha fatto!). Qualora al posto di 'che' l'introduttore fosse 'quale', la frase sarebbe meno accettabile a livello di italiano standard ('Quali sforzi che ha fatto')³⁸. Alla luce di quanto detto, dunque, il 'che' introduttore di una frase come 'Che sogno che ho fatto!' probabilmente non deve

³⁵ L. Renzi – G. Salvi – A. Cardinaletti ed., *Grande Grammatica Italiana di Consultazione*.

³⁶ Tratto da G. Berruto, *Sociolinguistica dell'italiano contemporaneo*.

³⁷ L. Renzi – G. Salvi – A. Cardinaletti ed., *Grande Grammatica Italiana di Consultazione*.

³⁸ Certo, l'impiego di vari introduttori di esclamativa è fortemente influenzato dalla diatopia. Inoltre, va rilevato che tali frasi potrebbero anche essere interpretate come interrogative di tipo retorico.

essere considerato polivalente. Esso chiaramente condivide alcune caratteristiche con i tipi di ‘che’ polivalente che tratterò in questo lavoro: si comporta variabilmente sul piano sociolinguistico, non è facilmente inquadrabile a livello semantico, mostra la coesistenza di proprietà funzionali distinte. Seppure questo ‘che’ in più ambiti neutralizzi delle differenze che sarebbero spesso esplicite in italiano standard, però, non sembra rientrare nella definizione di ‘che’ polivalente data in precedenza.

Si parla di ‘che’ polivalente anche quando ‘che’ sostituisce pronomi come ‘cui’ o ‘il quale’, preceduti dalla preposizione adatta, in frasi relative (si vedano i casi 16-18 della Tabella 1). Renzi, Salvi e Cardinaletti³⁹ individuano come elementi che possono essere relativizzati da ‘che’ soggetto, complemento oggetto, determinati complementi temporali (solo quelli che non richiedono la presenza di una preposizione: ‘il giorno che ti ho conosciuto portavi un cappotto di cammello’), sintagmi nominali che fungono da predicato nominale della copula, e alcuni complementi di misura. L’italiano standard prevede che il soggetto e l’oggetto diretto possano essere relativizzati per mezzo di ‘che’ (e.g. Il ragazzo che è partito era mio amico), mentre altre funzioni sintattiche devono essere relativizzate per mezzo di un pronome relativo ‘il quale’ o ‘cui’ accompagnato da una preposizione segnalante la funzione sintattica dell’elemento relativizzato (e.g. La ragazza con cui sono uscito è simpatica)⁴⁰. In quest’ultimo caso, però, le varietà sub-standard dell’italiano (così come sempre più l’italiano neo-standard) ammettono tre altre possibili costruzioni: la frase relativa può essere introdotta da ‘che’, senza alcuna marca della funzione sintattica dell’elemento relativizzato (e.g. Una busta che non so il contenuto); la frase relativa può essere introdotta da ‘che’, con la presenza di una marca della funzione sintattica dell’elemento relativizzato (e.g. I due americani che gli ho aperto l’ombrellone); la frase relativa può contenere una doppia codifica dell’elemento relativizzato (e.g. Ha ammonito la Francia a non compiere una nuova crociata in Libano, dalla quale ne uscirà sconfitta)⁴¹. Di queste tre possibilità, le prime due sono esempi di clausole subordinate che identificano/caratterizzano una testa nominale (e cioè dallo stesso valore funzionale delle frasi relative), introdotte dal solo complementatore ‘che’, alle quali Fiorentino fa riferimento con l’etichetta di ‘clausole relative deboli’⁴². Pertanto, simili occorrenze di *che* rientrano nei casi di polivalenza. Lo stesso discorso vale per costruzioni come ‘il giorno che ti ho conosciuto’.

³⁹ L. Renzi – G. Salvi – A. Cardinaletti ed., *Grande Grammatica Italiana di Consultazione*.

⁴⁰ M. Cerruti, *Costruzioni relative in italiano popolare*, in *Italiano e dialetto bresciano in racconti di partigiani*, F. Guerini ed., Aracne, Roma 2016.

⁴¹ È interessante notare come tali costruzioni siano largamente attestate anche nell’italiano antico. Per esempio, in Petrarca è impiegato il subordinatore invariabile ‘che’ per la relativizzazione di un locativo: “Questa vita terrena è quasi un prato / che l’ serpente tra’ fiori e l’erba giace” (*Canzoniere*, XCIX, vv. 5-6). L’uso anomalo di ‘che’ al posto di ‘nel quale’ in questo contesto è sottolineato anche nelle *Prose della Volgare Lingua* (1.III, Cap. LXIII).

⁴² G. Fiorentino, *Relativa debole*. Questo perché tali frasi sono introdotte da un complementatore invariabile che assolve a una generica funzione sul piano dell’enunciato, ovvero l’annunciare un’aggiunta di informazione rispetto a un tema, e non da un pronome (elemento linguistico dotato al contrario di capacità referenziali anaforiche forti).

Per quanto riguarda invece sintagmi nominali che fungono da predicato nominale della copula (Da gentiluomo che era rinunciò al suo diritto) e determinati complementi di misura (I tre metri che misurava erano davvero pochi), essi sono assimilabili a relativizzazioni di posizioni nominali non precedute da preposizione. Non li farò dunque rientrare nella polivalenza del 'che'.

Terminata quest'analisi sulla grammatica di Renzi, Salvi e Cardinaletti, si può riassumere quali siano le varie tipologie di 'che' polivalente che è possibile identificare sulla base della definizione operativa di polivalenza del 'che' fornita in precedenza (Tabella 1).

Tabella 1 - Tipologie di 'che' polivalente

<i>Valore</i>	<i>Esempio</i>
(11) Esplicativo-consecutivo	Vieni che ti pettino.
(12) Causale	Vai a dormire che ne hai bisogno.
(13) Consecutivo-presentativo	Io sono una donna tranquilla che sto in casa, lavoro*.
(14) Relativo temporale	Maledetto il giorno che ti ho incontrato.
(15) Pseudorelativo	Li vedo che scendono**.
(16) Sostituzione pronome	Ho sentito delle cose che al limite non avevo fatto caso.
(17) Sostituzione pronome	Il paese che sono stato domenica scorsa si chiama P.
(18) Sostituzione pronome	L'amico che stavo parlando un attimo fa è una vecchia conoscenza.

* Tratto da R. Sornicola, *Sul parlato*.

** Tratto da G. Berruto, *Sociolinguistica dell'italiano contemporaneo*.

Come già accennato, nella Tabella 1 le varie tipologie di 'che' polivalente sono definite su base semantico-sintattica, più che sociolinguistica. Questo significa che le differenze alla base delle varie tipologie del 'che' nella Tabella 1 non sono dovute a quanto vicine all'italiano standard le varie tipologie del 'che' polivalente possano sembrare (visto anche che cosa italiano standard sia tende a variare da individuo a individuo). I criteri che ho scelto di utilizzare per differenziarle vogliono rendere conto unicamente dei differenti contesti semantico-sintattici in cui i vari 'che' si possono trovare. Questa scelta è necessaria, visto il tipo di corpus su cui verrà condotto lo studio (annotato solo per *part of speech*).

Nei casi (11-15), i 'valori' associati a ciascun esempio sono dettati primariamente dalla semantica del tipo di relazione sintattica identificata dalla congiunzione, locuzione, ecc., alla quale il 'che' si sostituisce nei vari casi. Per esempio, considerando il caso (12) (Vai a dormire che ne hai bisogno.), che occorre in luogo di 'poiché' (Vai a dormire poiché ne hai bisogno); in tal senso, il valore attribuito a 'che' è causale. Questo discorso non è chiaramente valido nei casi (16-18), i quali si comportano in maniera diversa sul piano sintattico. È dunque ancora una volta evidente come all'unica categoria di 'che polivalente' debbano essere in realtà ricondotte due tipologie di fenomeni linguistici ben distinte: quella del 'che' complementatore impiegato come generico introduttore di frasi e quella del 'che' introduttore di frasi relative restrittive, le quali modificano un sintagma nominale. I che appartenenti alla seconda di queste categorie possono essere definiti restrittivi; i 'che' appartenenti alla prima di queste categorie possono essere definiti giustappositivi. Per

chiarire cosa s'intenda con 'che' restrittivi, può essere utile considerare il caso di alcune frasi relative in italiano introdotte da 'che'. Alcuni che relativi hanno infatti valore restrittivo nei confronti dell'elemento con cui sono in relazione. Possono essere definiti in questo modo perché fanno capo a una struttura di frase relativa la quale 'restringe' l'insieme dei referenti denotati dal sintagma nominale. Nella frase 'L'amico che ho visto ieri è molto alto', ad esempio, la frase relativa introdotta da 'che' restringe il numero di tutti i miei possibili amici, evocati dal sintagma nominale 'L'amico', al solo amico che ho visto ieri. In questo senso, anche alcuni dei tipi di 'che' polivalente visti possono avere valore restrittivo. Di contro, altri 'che' non hanno valore restrittivo (non restringono il significato dell'elemento cui si riferiscono), ma giustappositivo. La loro definizione come 'giustappositivi'⁴³ deriva dal fatto che giustappongono due proposizioni senza esplicitare un legame sintattico forte tra di esse (ma al limite solo una sfumatura a livello semantico).

Per chiarire con un esempio, inerente a casi di polivalenza del 'che', possono essere considerate la frase (17) e la frase (11):

(17) Il paese che sono stato domenica scorsa si chiama P.

(11) Vieni che ti pettino.

Nel primo caso, la frase introdotta da 'che' ('sono stato domenica scorsa') restringe il valore di 'paese', cui il 'che' è riferito. Nel secondo, invece, il 'che' non ha valore simile, ma di semplice giustapposizione di due proposizioni.

In riferimento alle tipologie di 'che' polivalente identificate nella Tabella 1, si possono 'etichettare' come giustappositivi i 'che' dei tipi (11), (12) e (15), mentre possono essere definiti restrittivi i 'che' dei tipi (14), (16), (17) e (18). Pochi dubbi emergono per i casi in (11) e (12). Il 'che' introduce una frase giustapposta alla precedente. Certo, la sfumatura semantica differisce, ma il valore non-restrittivo del 'che' appare chiaro. Parimenti chiaro, dall'altro lato, è il valore restrittivo del 'che' in contesti come il (14), il (16), il (17) o il (18). Di non facile identificazione è invece il tipo (13): 'Io sono una donna tranquilla che sto in casa, lavoro'. Esso è un caso particolare e non è facilmente precisabile a quale delle due categorie sia da ricondurre. Da un lato, infatti, il 'che' non è restrittivo di 'donna', come sarebbe se fosse relativo, e potrebbe essere inteso come congiunzione (Io sono una donna tranquilla e sto in casa, lavoro). Dall'altro, una debole referenza al soggetto (Io) non può essere negata. Probabilmente il suo esatto valore è da collocare nel *continuum* tra i due, senza una più precisa specificazione. Da ultimo, nel caso (15) (come verrà approfondito nella Sezione 4.4), il 'che' non si riferisce a 'li', ma ha valore giustappositivo (per intendersi, la frase potrebbe essere riscritta come: 'Li vedo mentre scendono'). In realtà, anche in questo caso è possibile rintracciare un debole valore restrittivo del 'che' (Vedo loro che (i quali) scendono). Ad ogni modo, il valore giustappositivo del 'che' è senza dubbio principale. Nello studio sul corpus verrà indagato se vi sia una correlazione o meno tra questa bipartizione delle tipologie dei 'che' e la loro rintracciabilità in un corpus taggato unicamente per *part of speech*.

⁴³ D. Wiechmann, *Understanding Relative Clauses. A Usage-Based View on the Processing of Complex Constructions*, De Gruyter, Berlin/New York 2015. Il termine inglese usato da Wiechmann, qui tradotto, è *appositive*.

3. Metodologia

La ricerca del 'che' polivalente in un corpus taggato unicamente per *part of speech* è difficoltosa. Il 'che' polivalente infatti non co-occorre (in maniera univoca) con una qualunque altra parte del discorso con un alto livello di probabilità.

In generale, la ricerca di determinate strutture linguistiche in un corpus mediante *query* si rivela efficace nel caso in cui alcune specifiche 'etichette' (es. DET, NOUN, o AUX) vengano con successo associate all'elemento lessicale corrispondente (es. 'il', 'cane', 'avere'). In questo modo, non vi è nessuna difficoltà a rintracciare in un corpus frasi come 'Rifletto su un libro', ricercando strutture sottostanti attraverso *query*, come [lemma="riflettere"] [tag="PRE"] [{"1"}] [tag="NOUN"]⁴⁴.

Le *query* sono uno strumento euristico efficace, e assumono valore ancora maggiore se combinate a parametri di tipo statistico. Con una *query* come [lemma="piovere"] [tag="NOUN" & lemma!="goccia|grandine"], ad esempio, è possibile rintracciare tutti gli usi figurati del verbo 'piovere'⁴⁵. Fondamentale a tal proposito è il dato statistico che indica quali siano i sostantivi più frequentemente presenti in posizione post-verbale rispetto a questo verbo⁴⁶. Tra di essi, 'goccia' e 'grandine' sono i più frequenti nel caso di usi non metaforici di 'piovere'. Questi lemmi sono quindi esclusi, nella posizione successiva al verbo, nella ricerca degli usi metaforici di 'piovere' con la *query* precedente ([tag="NOUN" & lemma!="goccia|grandine"]).

Tutto ciò con il 'che' polivalente non è possibile, e per ottenere risultati significativi la base statistica di co-occorrenza richiesta deve essere più raffinata. Questo può essere compiuto non solo ricercando quanto normalmente è presente nel contesto di un 'che' polivalente, ma anche escludendo tutto ciò che è probabilmente presente nel contesto di un 'che' di altra natura (ad esempio dichiarativo, vedi Sez. 4.1), e che lo identifica come tale.

Essendo il 'che' polivalente, come visto, un fenomeno di origine sociolinguisticamente 'non-standard', il corpus scelto per quest'analisi è composto da testi di scritto poco controllato (scritto dal web). Il corpus in esame è *ItWac*, il quale è creato tramite *web crawling*⁴⁷. Un efficace strumento per interrogare *ItWac*, in questo caso, è il *Corpus Query System Sketch Engine*⁴⁸. *ItWac* verrà interrogato tramite *query*. Queste, così come l'ana-

⁴⁴ Il formalismo impiegato qui (e in tutto lo studio) segue il *tagset* di *ItWac* (M. Baroni – S. Bernardini – A. Ferraresi – E. Zanchetta, *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora*, "Language Resources & Evaluation", 43, 2009, 3, pp. 209-226). La *query* qui presentata permette di rintracciare contesti in cui il lemma 'riflettere' è seguito da una preposizione, a sua volta seguita da un nome. Tra la preposizione e il nome può esservi al massimo un altro elemento lessicale (questo è indicato da [{"1"}]).

⁴⁵ La *query* qui presentata permette di rintracciare contesti in cui il lemma *piovere* è seguito da un nome diverso da 'goccia' o 'grandine' (questo è indicato da [tag="NOUN" & lemma!="goccia|grandine"]).

⁴⁶ In un *Corpus Query System* come *Sketch Engine*, questo dato può essere ottenuto utilizzando la funzione *word sketch* applicata al lemma 'piovere' e selezionando 'postV_N' come *gramrel*.

⁴⁷ Un *crawler* è un software che analizza i contenuti di siti web e database, creando generalmente una copia testuale dei documenti visitati.

⁴⁸ *Sketch Engine* è un *Corpus Query System* che permette all'utente di verificare i cosiddetti *sketch* di una parola (ovvero una lista dei possibili candidati per ognuna delle relazioni grammaticali connesse dalla parola in esame;

lisi qualitativa *in toto*, saranno composte a partire dal *Corpus Query Language* di *Sketch Engine*⁴⁹.

Per lavorare sui vari tipi di ‘che’ polivalente, agirò su due livelli in parallelo: da un lato quello delle *query* con cui interrogare il corpus, dall’altro quello di strutture semantico-sintattiche cui i singoli casi di ‘che’ polivalente sono riconducibili, costruite a partire dalla Tabella 1. Il principio alla base del procedimento che utilizzerò è il seguente: dopo aver ‘derivato’ una struttura semantico-sintattica per ogni tipologia di ‘che’ polivalente, se nel corpus verrà individuato un *pattern* contestuale in grado di ‘saturare’ tale struttura, il *che* all’interno di questo *pattern* sarà all’x% di probabilità polivalente. L’entità di questa percentuale verrà fornita dal corpus, interrogato tramite *query*.

Un’ulteriore precisazione necessaria prima di procedere con il lavoro sul corpus riguarda la *precision* (percentuale di accuratezza dei dati estratti) delle *query*⁵⁰. Obiettivo di questo lavoro non è trovare le *query* più raffinate possibili per ottenere una percentuale di accuratezza maggiore, ma verificare limiti e possibilità di un’estrazione tramite *query* di ‘che’ polivalenti all’interno di determinati *pattern* semantico-sintattici. Conseguentemente, le *query* che presenterò non avranno pretesa di essere esaustive e sicuramente sarebbero perfezionabili. Esse andranno piuttosto in direzione di un’ottimizzazione dei risultati. Spesso infatti ricercherò contesti estremamente specifici, i quali, seppure causando la perdita di una parte di risultati utili, permetteranno di ridurre in maniera considerevole il rumore dovuto a risultati che non mostrano la tipologia di ‘che’ oggetto di ricerca. Per il medesimo motivo (ovvero ridurre per quanto possibile il ‘rumore’) ho scelto di utilizzare un corpus come *ItWac* e non altri corpora dalle stesse caratteristiche ma più rappresentativi (es. *itTenTen10*).

Infine, è necessario approfondire la decisione di non valutare la *recall*⁵¹ (percentuale dei dati-*target* estratti sul totale dei dati-*target* presenti nel corpus) delle *query* che impiegherò⁵². Non compirò questa operazione principalmente per due motivi. In primo luogo, perché – esattamente come nel caso della *precision* – non è scopo di questo lavoro individuare nella misura migliore possibile tutti i tipi di ‘che’ polivalente in un corpus POS-tagato (e dunque il massimo grado di accuratezza nei termini di *precision* e *recall* delle *query* nei vari casi). In secondo luogo, la creazione di un *gold standard* significativo per il ‘che’ polivalente (ovvero un estratto del corpus contenente alcuni casi di ‘che’ polivalente correttamente annotati come tali) è una questione che pone non pochi problemi vista la bassissima frequenza

per esempio, nel caso di un verbo, queste sono: il soggetto, l’oggetto, altri verbi che frequentemente co-occorrono con il verbo in questione, avverbi, ecc.), differenze tra *sketch* di parole diverse, in aggiunta alle più classiche funzioni di CQS. Attraverso *Sketch Engine* sono interrogabili numerosissimi corpora, in decine di lingue del mondo.

⁴⁹ <http://www.sketchengine.co.uk/documentation/wiki/SkE/CorpusQuerying#CorpusQueryLanguageC-QL> ultima consultazione 27 agosto 2016.

⁵⁰ C.J. Van Rijsbergen, *Information Retrieval*, Butterworths, London 1979.

⁵¹ *Ibidem*.

⁵² In uno studio basato su corpora è necessario valutare non solo la *precision* delle *query* utilizzate per estrarre i dati dal corpus, ma anche la loro *recall*. Se tale operazione non viene compiuta, si corre il rischio di individuare tramite *query* dati estremamente precisi, ma di ignorarne una grandissima quantità. La valutazione della *recall* delle *query* deve avvenire su un *gold standard* estratto randomicamente dal corpus.

del numero di 'che' polivalenti sul numero di 'che' totali presenti nel corpus. Qualora il *gold standard* fosse creato come *sample* di un estratto randomico di *ItWac*, le probabilità di avere 'che' polivalenti sarebbero pressoché nulle. Allo stesso modo, vicine allo zero sarebbero le probabilità di trovare 'che' polivalenti in un *gold standard* creato come *sample* di un estratto randomico di *ItWac* in cui la presenza di un 'che' è imposta come obbligatoria (attraverso [tag="CHE"]). Inoltre, in questo secondo caso sarebbe complesso verificare la *recall* della *query* utilizzata per cercare nel corpus il 'che' polivalente di un tipo specifico, vista la differenziazione di varie tipologie di 'che' polivalente su cui lavorerò.

Una possibile soluzione potrebbe essere quella di creare il *gold standard* a partire dalle singole *query* impiegate per estrarre i vari tipi di 'che' polivalente. Le tipologie di 'che' polivalente individuate nella Sezione 2 sono sei (esplicativo-consecutivo, causale, consecutivo-presentativo, relativo temporale, pseudorelativo, sostituzione pronomi). Lanciate le *query* in *ItWac*, il *gold standard* potrebbe essere creato sommando il 16,6% (1/6) dei risultati selezionati randomicamente tra i dati estratti dal corpus per ognuna delle tipologie di *che* polivalente, poi annotati manualmente sulla tipologia del *che* (di ogni *che* andrebbe specificato se relativo, completivo, polivalente del tipo a, polivalente del tipo b, ecc.). Tuttavia, il rischio di circolarità è elevato, e il *gold standard* creato in questo modo sarebbe probabilmente non adeguatamente rappresentativo del corpus. Esso sarebbe infatti pesantemente influenzato dalla forma della *query*, la quale rischierebbe di limitare l'universo di riferimento in maniera troppo grossolana.

Come è chiaro, quindi, la questione non è di facile soluzione: se si vuole preservare la rappresentatività del *gold standard*, si corre il rischio di non avere abbastanza dati vista la bassa frequenza dei 'che' polivalenti sul totale dei 'che' presenti nel corpus; se si vuole esser certi di avere casi sufficienti per ogni tipologia di 'che' polivalente, è verosimile che la rappresentatività del *gold standard* sia compromessa.

Probabilmente, l'unico modo in cui può essere creato un *gold standard* efficace è estrarre dal corpus *ItWac* un numero molto elevato di contesti in cui sia presente un 'che' e annotarli manualmente uno per uno, rinunciando a un'operazione di *sample*. Tuttavia, quest'analisi non verrà effettuata. Visti i fini di questo lavoro, per avere un termine di paragone sui dati estratti dal corpus con le singole *query* mi limiterò a verificare i risultati che si ottengono con le stesse *query* in un altro corpus dalle stesse caratteristiche di *ItWac*, ovvero *itTenTen10*, il quale condivide con *ItWac* la possibilità di essere interrogato attraverso il *Corpus Query Language* di *Sketch Engine* ed è POS-tagato secondo lo stesso *tagset*⁵³. Qualora le evidenze trovate fossero assimilabili, esse sarebbero uno spunto di riflessione sufficientemente valido.

⁵³ M. Baroni – S. Bernardini – A. Ferraresi – E. Zanchetta, *The WaCky Wide Web*.

4. *Analisi quantitativa*

4.1. *'Che' esplicativo-consecutivo (11) e 'che' causale (12)*

I 'che' polivalenti di tipo esplicativo-consecutivo (11) (Vieni che ti pettino.) e causale (12) (Vai a dormire che ne hai bisogno.) esprimono un legame tra due azioni. In realtà, le sfumature semantiche del 'che' riconducibili a questa tipologia non sono solamente esplicativo-consecutive o causali. Un esempio di ciò è un 'che' polivalente dal valore che potrebbe essere definito 'eccettuativo'⁵⁴. Apparentemente, questa e altre tipologie simili non vengono considerate in quest'analisi. Una simile scelta è dettata dal fatto che in tali casi il *pattern* sintattico, a fronte di una variazione semantica, rimane lo stesso ad esempio del caso (11). Conseguentemente, gli uni e l'altro possono essere assimilati. Semplificando, si può dire che entrambi (11) e (12) siano riconducibili a una struttura del tipo: V_1 (verbo) – che – V_2 .

La stessa struttura V_1 – che – V_2 è alla base anche di 'che' di tipo dichiarativo o completivo (i quali ovviamente non rientrano nei casi di polivalenza del 'che'). Per identificare su base contestuale il 'che' esplicativo-consecutivo e quello causale, dunque, l'individuazione di un contesto che rientri in V_1 – che – V_2 non è sufficiente. Per iniziare a raffinare la ricerca, si può pensare di eliminare dal contesto di 'che' tutti quei verbi che reggono una frase dichiarativa o completiva introdotta da 'che'. Per ricercarli, è sufficiente lanciare una ricerca con una *query* come la (19):

(19) [tag="VER.*"][tag="CHE"][tag="VER.*"]⁵⁵.

In questo modo si trovano contesti come il (20), i quali erano esattamente il target della *query*:

(20) *credo che* sia l'amministrazione, sia il collegio si siano mossi [...].

Tuttavia, identificati dalla *query* (19) sono anche contesti simili al (21):

(21) Gesù crocifisso e *risorto che vive* trionfante sul peccato [...].

Il 'che' in questione è di tipo relativo e aderisce alla norma dell'italiano standard. A sua volta, dunque, questa tipologia di 'che' va esclusa ai fini della ricerca di (11) e (12). L'ambiguità nasce dal tag VER dato ai participi passati. Per aumentare il grado di precisione della ricerca, quindi, si può pensare di eliminare dal contesto sinistro del 'che' i participi passati, con una *query* come la (22):

(22) [tag="VER.*"&tag!="VER:ppast"][tag="CHE"][tag="VER.*"].

⁵⁴ Non può che mangiare verdura.

⁵⁵ Questa *query* permette di individuare frasi in cui un verbo è seguito da 'che', il quale a sua volta è seguito da un altro verbo (e.g. Luigi dice che mancano due ore).

Fatto ciò, si ottengono risultati per oltre 770.000 occorrenze. Per poter escludere in maniera statisticamente efficace tutti quei verbi che reggono un 'che' di tipo dichiarativo o completivo, può essere effettuato un *sample* a 250 occorrenze. In questo modo, controllando uno per uno i contesti, è possibile stilare una lista dei verbi che introducono una dichiarativa o una completiva. Essi sono elencati in (23).

- (23) Ricordare, dimostrare, raccontare, parere, sembrare, sentire, dire, confidare, convenire, sperare, vedere, volere, sapere, scrivere, rispondere, scoprire, ribadire, promettere, mettere, insistere, dichiarare, osservare, sostenere, immaginare, temere, ritenere, premettere, augurare, osservare, pensare, pretendere, calcolare, lasciare, ripetere, scommettere, riconoscere, significare, vedere, prevedere, constatare, supporre, importare, capire, trovare, decidere, precisare, aggiungere, ammettere, notare, credere, chiedere, aspettare, considerare, bisognare, occorrere, avvenire, bastare, essere.

Seppure si trovino molti errori di *tagging*, come il (24), e seppure il metodo in questione possa apparire anti-economico, può comunque essere considerato abbastanza efficace.

- (24) Vi è poi la rete *ANPA che stiamo* costituendo [...].

Questo processo permette di giungere a una *query* come la (25), volta – in definitiva – a cercare contesti dalla struttura $V_1 - che - V_2$, in cui V_1 non sia un participio passato o uno dei verbi elencati in (23):

- (25) [tag="VER.*"&tag!="VER:ppast"&lemma!="ricordare | dimostrare | raccontare | parere | sembrare | sentire | dire | confidare | convenire | sperare | vedere | volere | sapere | scrivere | rispondere | scoprire | ribadire | promettere | mettere | insistere | dichiarare | osservare | sostenere | immaginare | temere | ritenere | premettere | augurare | osservare | pensare | pretendere | calcolare | lasciare | ripetere | scommettere | riconoscere | significare | vedere | prevedere | constatare | supporre | importare | capire | trovare | decidere | precisare | aggiungere | ammettere | notare | credere | chiedere | aspettare | considerare | bisognare | occorrere | avvenire | bastare | essere"] [tag="CHE"] [tag="VER.*"].

Il numero di occorrenze trovate è in questo modo ridotto a meno di 300.000. Sempre per facilità di lavoro, si può operare un *sample* a 250 contesti. Dall'analisi dei risultati così ottenuti, si evidenziano ancora numerosi errori di *tagging* (26), casi di 'che' completivo o dichiarativo sfuggiti alla (25) perché retti da verbi non presenti nel *sample* precedentemente effettuato (27)⁵⁶, casi di 'che' relativo aderenti allo standard dovuti a errori di *tagging* (28).

⁵⁶ Un *sample* è eseguito in maniera funzionale all'elaborazione di dati in numero così elevato. Esso implica necessariamente una generalizzazione, con conseguente perdita di dati. Non sarebbe stato pensabile controllare una a una le oltre 770.000 occorrenze identificate in precedenza. Bisogna in questi casi prendere atto dell'esi-

- (26) *pellegrini che desiderano* rivivere le emozioni provate alle Catacombe [...].
 (27) *Giurano che va matto* per la Nutella.
 (28) [...] provocando *attese che hanno* pesanti ripercussioni [...].

Ad ogni modo, si trovano anche contesti come il (29), il quale presenta polivalenza del ‘che’:

- (29) *beve che sembra* un acquaio [...].

Come visto, gli errori di *tagging* sono molti, e così anche i verbi sfuggiti alla sommaria lista in (23). Tuttavia, rintracciare una struttura sintattica – per quanto complessa – alla base di ‘che’ polivalenti di tipo esplicativo-consecutivo e causale appare possibile. Essa può essere schematizzata con la (30):

- (30) $V_1 - \text{che} - V_2$, dove V_1 non è un participio passato o uno dei verbi in (23).

Certo, l’altissimo grado di imprecisione del POS-*tagging* in *ItWac* rende difficilmente quantificabile in termini precisi con quale probabilità in percentuale un contesto individuato tramite *query* aderente alla struttura (30) manifesti una polivalenza del ‘che’. In un corpus però maggiormente controllato da questo punto di vista e con l’aiuto di una risorsa lessicale per la completezza della (23), il grado di precisione della struttura vista in (30) per rendere conto di ‘che’ esplicativo-consecutivi o causali è potenzialmente soddisfacente.

4.2. ‘Che’ consecutivo-presentativo (13)

È possibile ricondurre il *che* polivalente di tipo consecutivo-presentativo (e.g. Io sono una donna tranquilla che sto in casa, lavoro.) a una struttura di massima del tipo: S (soggetto) – C (copula) – Ap (apposizione) – che – V, dove il ‘che’ ha valore giustappositivo e non restrittivo. Tuttavia, già a prima vista è chiaro come ancora una volta questa struttura sia troppo generale⁵⁷.

In primo luogo è necessaria un’identità di soggetto tra C e V affinché il ‘che’ in questione sia consecutivo-presentativo. Ciò è verificabile riprendendo e modificando l’esempio in (13). Se non vi fosse identità di soggetto, infatti, non avremmo altro che casi di ‘che’ relativo standard: ‘Tu sei una donna tranquilla che sta in casa’. In secondo luogo, altro elemento imprescindibile è che S non sia di terza persona. Anche in questo caso il *che* sarebbe relativo aderente allo standard, o perlomeno sarebbe impossibile distinguere se relativo o

stenza di altri verbi che reggono un ‘che’ di tipo dichiarativo o completivo (mancano, ad esempio, verbi come ‘desiderare’), e di volta in volta aggiungerli alla lista in (23). Una lista completa può comunque essere stilata sulla base di un dizionario o di una risorsa lessicale.

⁵⁷ La *query* che più direttamente rappresenta tale struttura sarebbe simile a [lemma=“essere”][tag=“NOUN”][tag=“CHE”][tag=“VER.*”]. Una *query* del genere è la stessa che permette di individuare ‘che’ di tipo relativo aderenti allo standard (e.g. [...] sono furgoni che trasportano 15, 20 persone [...]).

consecutivo-presentativo da un punto di vista sintattico: Maria è una donna tranquilla che sta in casa.

Ecco quindi che la struttura sottostante al 'che' consecutivo-presentativo può essere riconosciuta in: S (soggetto) – C (copula) – Ap (apposizione) – che – V, dove S non è di terza persona, e dove C e V hanno lo stesso soggetto S. Rintracciare su base contestuale una struttura del genere in un corpus POS-taginato è impossibile.

Ad ogni modo, questo tipo di 'che' polivalente non è da abbandonare così rapidamente. A esso, infatti, possono essere ricondotti contesti come il (31). In entrambi i casi, infatti, il 'che' si comporta da generico introduttore di frase (dalla semantica non chiaramente specificata) con valore giustappositivo.

(31) *Che magari vai* in Francia, in Canada, in Australia [...].

In generale, questo *che* è sempre preceduto da un segno di punteggiatura, il quale può essere forte (come un punto o un punto esclamativo, [tag="SENT"]) oppure debole (come una virgola, [tag="PUN"]). La struttura sottostante a frasi come la (31) è semplice: (punteggiatura) che – V. Tuttavia, ogni volta in cui un 'che' segue un segno di punteggiatura, specialmente se debole, è altamente probabile che sia relativo in maniera aderente all'italiano standard. Tutto ciò è rapidamente verificabile con una *query* come la (32), la quale restituisce principalmente contesti assimilabili al (33).

(32) [tag="PUN"][tag="CHE"][]{}{0,2}[tag="VER.*"]⁵⁸,

(33) [...] su apposito registro, *che dovrà* essere esibito in visione [...].

Di maggiore interesse è il caso della *query* (34). Essa considera solo segni di punteggiatura forte, abbassando in teoria il 'rischio' di incontrare 'che' relativi standard:

(34) [tag="SENT"][tag="CHE"][]{}{0,2}[tag="VER.*"].

Lanciando la ricerca, si ottengono oltre 220.000 contesti in risultato. Effettuando un *sample* a 250 occorrenze, si può notare come il 'che' sia polivalente in pochissimi casi (e.g. Che non si trattasse solo dell'abituale reazione [...]). Tutto ciò porterebbe a pensare che anche per il 'che' polivalente esemplificato in (31) un riconoscimento su base contestuale non sia possibile. In realtà, la percentuale di polivalenza del 'che' rintracciabile con la (34) sale vertiginosamente se, nello *slot* immediatamente successivo a [tag="CHE"], vengono ricercati alcuni tipi di avverbi, la maggior parte dei quali appartiene al gruppo degli avverbi di dubbio. La (34) può quindi essere riformulata nella (35):

(35) [tag="SENT"][tag="CHE"][word="se | forse | chissà | eventualmente | probabilmente | poi | magari"][]{}{0,1}[tag="VER.*"];

⁵⁸ Questa *query* permette di identificare frasi in cui un *che* segue un segno di punteggiatura forte. Dopo il 'che', separato al massimo da due parole ([[]{}{0,2}), occorre un verbo.

La ricerca rende più di 4.600 occorrenze come risultato. Effettuando un *sample* a 250 contesti è possibile vedere come, grossomodo, il numero di ‘che’ polivalenti sia oltre il 90% del totale (e.g. Che forse non trova quello che cerca perché non sa bene cosa sta cercando?). Un contesto del tipo (punteggiatura forte) che – AdD (avverbio di dubbio) – V, dunque, prevede la polivalenza del ‘che’ in più di nove casi su dieci.

4.3. ‘Che’ relativo-temporale (14)

Approfondirò il discorso relativo al ‘che’ polivalente di tipo relativo-temporale nella Sezione 4.5. Esso presenta, infatti, caratteristiche che lo assimilano ai casi di uso generalizzato del ‘che’ nelle frasi relative in sostituzione di un pronome relativo obliquo ‘cui’ o ‘il quale’.

4.4. ‘Che’ pseudorelativo (15)

Una frase come la (15) (Li vedo che scendono.) è molto utile per individuare la struttura all’interno della quale potrebbe essere rintracciato il *che* polivalente pseudorelativo⁵⁹. Essa potrebbe essere la seguente: Pn (pronome) – V₁ – che – V₂, dove V₂ è concordato con Pn. Se anche il ‘che’ sembra svolgere funzione di soggetto relativo di V₂, il suo valore è primariamente giustappositivo (proprio in ragione di ciò è detto pseudorelativo).

La *query* che più strettamente offre contesti simili al (15) è la (36):

(36) [word=“li”][word=“vedo|sento”][tag=“CHE”][tag=“VER.*”].

Certo, il ventaglio di occorrenze da essa estraibili è limitato (emblematico è il risultato di 12 contesti in *ItWac*; e.g. Dopo li sento che parlano.). Un primo ampliamento può essere fatto includendo altri pronomi in posizione preverbiale e sostituendo a [word=“vedo|sento”] i rispettivi lemmi. La (36) diviene quindi la (37):

(37) [word=“(m|t|c|v|l)(i’)|(l)(a|o)”][lemma=“vedere|sentire”][tag=“CHE”][tag=“VER.*”].

In questo modo si ottengono 780 contesti come risultato. Tuttavia, frasi come la (38) non rientrano nella struttura Pn (pronome) – V₁ – che – V₂, dove V₂ è concordato con Pn, pur corrispondendo alla *query* in (37). In un corpus solamente POS-tagato, il fatto che V₂ non debba essere concordato con Pn non è formalizzabile in una *query*. Dato che può essere considerato solo l’ordine sequenziale e lineare dei *tag*, accordo, dipendenza o referenza non possono infatti essere resi.

⁵⁹ In realtà, all’interno di questa categoria rientrano anche casi realizzati linguisticamente in maniera diversa. Si consideri ad esempio la seguente frase: ‘Ho sentito i ragazzi che scendevano’. Qui il complemento oggetto del verbo della principale non è pronominale, ma nominale (e dunque la struttura Pn – V₁ – *che* – V₂, dove V₂ è concordato con Pn sopra descritta non è valida). Tuttavia il tipo di polivalenza del ‘che’ è il medesimo della frase ‘Li vedo che scendono’. La differenza esistente tra queste due frasi non è però determinante ai fini dell’analisi compiuta in questo lavoro.

(38) Non *lo vedi che stiamo* lavorando?

Per escludere questa e occorrenze simili, ho scelto di eliminare 'lo' dall'elenco dei pronomi-complemento oggetto. Sicuramente questo ha causato la perdita di una certa quantità di dati; essi dovrebbero però essere più precisi. Un'altra operazione che può essere compiuta per aumentare il grado di generalizzazione della (37) è arricchire il [lemma="vedere|sentire"] con gli altri principali *verba sentiendi*. Essi sono elencati in (39):

(39) Vedere, guardare, scorgere, notare, osservare, mirare, sentire, ascoltare, udire, intendere, percepire.

La (37) diviene allora la (40):

(40) [word="(m|t|c|v|l)(i')|la"] [lemma="vedere | sentire | guardare | scorgere | notare | osservare | mirare | ascoltare | udire | intendere | percepire"] [tag="CHE"] [tag="VER.*"].

In questo modo, si ottengono 241 occorrenze. Oltre il 70% di esse rientra in Pn (pronome) – V₁ – che – V₂, dove Pn non è 'lo', V₁ è un *verba sentiendi*, V₂ è concordato con Pn (e.g. *Li vediamo che vanno ad Ocosingo [...]*).

Tuttavia, anche frasi come la (41) o la (42), le quali non rientrano nella struttura appena presentata, sono individuate dalla (40):

(41) [...] non è da chi *li nota che può* venire la soluzione [...].

(42) [...] quando *li vedi che fai*, ti tocchi la testa?

Esse però sono in numero minoritario. In generale, infatti, se un 'che' rientra in un *pattern* contestuale individuato dalla *query* (40), con oltre il 70% di probabilità è polivalente.

4.5. Sostituzione di pronome (16-18)

La struttura semantico-sintattica cui ricondurre i casi (16) (Ho sentito delle cose che al limite non avevo fatto caso.), (17) (Il paese che sono stato domenica scorsa si chiama P.) e (18) (L'amico che stavo parlando un attimo fa è una vecchia conoscenza.) è la seguente: N – che – V, dove V non è concordato con N o con 'che'. Conseguentemente, la *query* in grado di renderla più direttamente è la (43):

(43) [tag="NOUN"] [tag="CHE"] [tag="VER.*"].

Le occorrenze in *ItWac* sono più di 3.700.000. È sufficiente operare un *sample* a 250 contesti per rendersi conto di come la quasi totalità dei 'che' individuati non appartenga a casi di polivalenza. Le strutture che aderiscono alla (43) presentano praticamente tutte 'che' di tipo relativo aderenti allo standard (44):

- (44) [...] Ovvero la *legge che punisce* tutti coloro che vendono prodotti non originali [...].

La sostituzione di pronomi come ‘cui’ o ‘il quale’ tramite ‘che’ non è individuabile solo su base contestuale. Come già visto nell’analisi del tipo (13), infatti, vi è un problema di referenza che in un corpus solamente POS-tagato non può essere resa. Per chiarire il discorso, il fenomeno può essere analizzato a livello di valenza del verbo. Come già sottolineato, se un verbo è trivalente (es. dare), il ‘che’ relativo può assumere in italiano standard il ruolo di S di V (45) o di O (oggetto) di V (46).

- (45) L’amico che dava lavoro ai figli.
 (46) Il lavoro che l’amico dava ai figli.

Di contro, nel caso in cui ‘che’ assumesse il valore – in quanto argomento del verbo – di oggetto indiretto, il ‘che’ sarebbe polivalente (47).

- (47) I figli che l’amico dava il lavoro.

Per ricostruire tutto questo su base contestuale, sarebbe necessario un *tagging* in grado di assegnare gli argomenti verbali a V (oppure, chiaramente, la referenza del ‘che’). Ma questo non è previsto nel caso di *ItWac* o di un qualunque corpus solamente POS-tagato.

Lo stesso vale ovviamente per verbi bivalenti. V deve però essere caratterizzato dalla reggenza di un oggetto indiretto anziché diretto (es. pensare). In questo caso, il discorso può essere riproposto: se il ‘che’ è referente dell’oggetto indiretto (ovvero se il verbo è contestualmente saturato nel suo primo argomento)⁶⁰, allora è polivalente (48). In questa categoria rientrano anche casi come il (17) (Il paese che sono stato domenica scorsa si chiama P.). I verbi di movimento possono essere, infatti, intransitivi bivalenti quando telici⁶¹. Essi oltre al soggetto richiedono un argomento collegato al verbo per mezzo di una preposizione, definito ‘argomento indiretto’ o ‘preposizionale’.

- (48) Il libro che pensa Giovanni è avvincente.

In realtà, questo discorso sulla valenza verbale non rende conto della totalità dei casi. Un esempio valido può essere il (49):

- (49) Il coltello che taglio il pane.

⁶⁰ Questo discorso non vale per casi come: ‘La vita che penso essere triste’.

⁶¹ Questo può essere chiarito considerando ad esempio ‘correre’. Nel caso in cui ‘correre’ sia da intendersi come genericamente di attività esso è atelico e monovalente (e.g. Ho corso a perdifiato), nel caso in cui invece indichi moto direzionale esso è telico e necessita di una specificazione locativa per completarsi (e.g. Sono corso a casa).

'Tagliare' è bivalente e regge casi diretti. Nonostante ciò, in (49) il 'che' è polivalente e in sostituzione di un pronome relativo obliquo.

Nella Sezione 4.3 è stata rimandata a questa la spiegazione del 'che' relativo temporale, in quanto anche contesti come il (14) (Maledetto il giorno che ti ho incontrato.) sono riconducibili alla stessa struttura astratta N – che – V, dove V non è concordato con N o con 'che' già valida per i casi in (16), (17) o (18), con la differenza che N indica referenza temporale. Ovviamente i problemi sono i medesimi. Questo è facilmente verificabile attraverso una *query* come [lemma="giorno"][tag="CHE"][tag="VER.*"] (seppure essa non sia particolarmente raffinata). In sostanza, tutti gli oltre 16.000 contesti che si ottengono in risultato mostrano usi di 'che' di tipo relativo aderenti allo standard (Il 21 maggio, giorno che commemora il ritrovamento [...]), completivo (Accadde un giorno che gli giungesse voce dell'esistenza [...]) o propriamente temporale (Sono cinque giorni che non ho notizie di Massimo). Solo il 5% dei risultati rientra sotto questo tipo di polivalenza del 'che' (Ci sarà quel giorno che ti becchi la ricciola da 40 kg)⁶².

5. *Discussione*

La Tabella 2 riassume i tipi di 'che' polivalente che sembrano essere individuabili su base contestuale in un corpus POS-taginato e quelli che invece non sembrano rintracciabili. I casi in cui non è possibile individuare un *pattern* semantico-sintattico (che include un 'che' polivalente) estraibile con una determinata probabilità tramite *query* da un corpus POS-taginato come *ItWac* sono quelli in cui è presente una referenza testuale che va al di là del semplice ordine lineare delle parole. Conseguentemente, utilizzando uno strumento euristico (le *query*) in grado di rendere solo dipendenze di tipo orizzontale e sequenziale, tali strutture non possono essere rese. Ovviamente, il discorso è valido solo per corpora che presentano come unico livello di annotazione quello per *part of speech*.

Tabella 2 - Tipologie di 'che' polivalente rintracciabili e non rintracciabili in corpus annotati unicamente per *part of speech*

<i>'Che' rintracciabili in corpus POS-taginati</i>	<i>'Che' non rintracciabili in corpus POS-taginati</i>
(11) Esplicativo-consecutivo	(13) Consecutivo-presentativo*
(12) Causale	(14) Relativo-temporale**
(13) Consecutivo-presentativo***	(16-18) Sostituzione di pronome
(15) Pseudorelativo	

* Per contesti riconducibili al (13): Io sono una donna tranquilla che sto in casa, lavoro.

** Il 'che' polivalente relativo-temporale è inserito tra le tipologie di 'che' polivalente non rintracciabili su base contestuale nonostante occorra nel 5% dei casi totali in *ItWac* (e, come chiarito in seguito dalla Tabella 3, in circa il 10% dei casi totali in *itTenTen10*). Sebbene queste percentuali siano significative, va ricordato che la *query* che è stata utilizzata per rintracciarlo ([lemma="giorno"][tag="CHE"][tag="VER.*"]) è estremamente rigida. Essa esclude infatti casi del tutto analoghi ma con espressioni temporali differenti da 'giorno' (es. Maledetto

⁶² Tali proporzioni emergono grazie a uno studio dell'output della *query* tramite *sample* a 250 contesti.

il momento/mese/anno che ti ho incontrato/pensato). Qualora queste fossero state inserite come variabili, la percentuale di casi di ‘che’ relativo-temporale sui contesti totali con che estratti dal corpus sarebbe scesa sensibilmente. Aumentando le variabili sarebbe cresciuto, parallelamente, anche il rumore.

*** Solo per contesti riconducibili al (31): Che magari vai in Francia, in Canada, in Australia [...].

Per compiere una rapida valutazione dell’affidabilità di questi dati verificherò ora quali siano i risultati che si ottengono con le stesse *query* in *itTenTen10*, ovvero un corpus dalle stesse caratteristiche di *ItWac*. Il confronto tra i due corpora è presentato nella Tabella 3.

Tabella 3 - confronto tra *ItWac* e *ItTenTen10*. Per ogni tipologia di che polivalente sono riportati tre valori: il numero di risultati dati in output dalla *query*, la possibilità o meno di rintracciare il dato tipo di c’he’ polivalente nel corpus, la percentuale del numero di contesti che mostrano il ‘che’ del dato tipo sul numero dei contesti totali in output dalla *query*, valutati tramite un *sample* a 250 contesti

<i>Tipologia</i>	<i>ItWac</i>	<i>ItTenTen10</i>
(11) Esplicativo-consecutivo	~300.000/Sì/ND*	~535.000/Sì/ND**
(12) Causale	~35.000/No/~0%	~70.000/No/~0%
(13) Consecutivo-presentativo	~5.000/Sì/~90%	~11.000/Sì/~80%
(31) Consecutivo-presentativo	~16.000/No/~5%	~19.000/No/~10%
(14) Relativo-temporale	~250/Sì/70%	~500/Sì/~85%
(15) Pseudorelativo	~3.700.000/No/~0%	~7.000.000/No/~0%
(16-18) Sostituzione di pronomi		

* ND sta per ‘non determinato’.

** Questo dato non è determinabile sulla base della *query* in

(25) ([tag=“VER.”&tag!=“VER:ppast”&lemma!=“ricordare | dimostrare | raccontare | parere | sembrare | sentire | dire | confidare | convenire | sperare | vedere | volere | sapere | scrivere | rispondere | scoprire | ribadire | promettere | mettere | insistere | dichiarare | osservare | sostenere | immaginare | temere | ritenere | premettere | augurare | osservare | pensare | pretendere | calcolare | lasciare | ripetere | scommettere | riconoscere | significare | vedere | prevedere | constatare | supportare | importare | capire | trovare | decidere | precisare | aggiungere | ammettere | notare | credere | chiedere | aspettare | considerare | bisognare | occorrere | avvenire | bastare | essere”] [tag=“CHE”][tag=“VER.*”]) per le stesse motivazioni già viste nella Sezione 4.1

Come emerge dalla Tabella 3, i dati provenienti da *ItWac* e i dati provenienti da *ItTenTen10* sono del tutto assimilabili, in quanto l’individuabilità o meno delle varie tipologie di *che* polivalente risulta coerente nei due diversi corpora annotati per *part of speech* secondo lo stesso *tagset*. Le uniche differenze riguardano il numero di risultati per tipologia di *che* polivalente e la percentuale di casi con *che* polivalente di un dato tipo sul numero totale di contesti in output dalla *query*. La prima differenza è dovuta alle diverse dimensioni dei due corpora, la seconda alla loro differente composizione (le differenze non sono, ad ogni modo, di entità rilevante).

Può a mio avviso essere poi interessante tornare sulla distinzione introdotta nella Sezione 2 tra ‘che’ restrittivi e ‘che’ giustappositivi. È infatti degno di nota (Tabella 4) come – grossomodo – i casi di polivalenza del ‘che’ di tipo restrittivo siano quelli che nella Tabella 2 sono stati individuati come non rintracciabili su base contestuale; quelli di tipo giustappositivo, invece, paiono rintracciabili contestualmente.

Tabella 4 - classificazione dei sotto-tipi di 'che' polivalente in 'che' giustappositivi e in 'che' restrittivi. 'Sì' e 'No' indicano la rintracciabilità o meno di una data tipologia in un corpus POS-taginato

<i>'Che' giustappositivi</i>	<i>'Che' restrittivi</i>
(11) Vieni che ti pettino.	Sì (14) Maledetto il giorno che ti ho incontrato. No
(12) Vai a dormire che ne hai bisogno.	Sì (16) Ho sentito delle cose che al limite non avevo fatto caso. No
(31) Che magari vai in Francia, in Canada, in Australia [...].	Sì (17) Il paese che sono stato domenica scorsa si chiama P. No
(15) Li vedo che scendono.	Sì (18) L'amico che stavo parlando un attimo fa è una vecchia conoscenza. No

Nella Tabella 4, alla suddivisione in *che* restrittivi e giustappositivi operata nella Sezione 2 è stato aggiunto il tipo (31). Il 'che' occorrente in questo genere di frasi è certamente un 'che' generico introduttore di frase (e dunque giustappositivo). Non si può infatti dire che si riferisca a un elemento specifico restringendone la semantica. Inoltre, va precisato che il tipo (13) (e.g. Io sono una donna tranquilla che sto in casa, lavoro), precedentemente escluso dalla bipartizione tra 'che' restrittivi e 'che' giustappositivi e conseguentemente escluso dalla Tabella 4, non risulta individuabile su base meramente contestuale.

Questo interessante risultato mette in luce ancora una volta come il 'che' polivalente non sia un fenomeno linguistico unitario. Al contrario, all'interno di un'unica categoria rientrano disparati fenomeni linguistici, relativi a diversi livelli d'analisi. Si può tuttavia concludere che quando il 'che' polivalente (complementatore) viene impiegato come generico introduttore di frase (e.g. Sbrigati che è tardi!), anche in strutture in cui la dipendenza sintattica viene costruita a livello più testuale che frasale (e.g. Che poi magari ci vanno lo stesso!), esso è rintracciabile in un corpus facendo unicamente affidamento sulla sequenza lineare dei *tag* per *part of speech* del suo contesto. Al contrario, quando il 'che' introduce frasi relative restrittive, che cioè modificano sintagmi nominali (e.g. Maledetto il giorno che ti ho incontrato), esso non pare rintracciabile.

6. Conclusioni e prospettive future

In questo lavoro ho cercato di fare il punto della situazione sulla possibilità di individuare il 'che' polivalente (o meglio, i 'che' polivalenti) in un corpus taginato solamente per *part of speech*. Dopo una premessa di tipo sociolinguistico alla quale è seguita un'analisi qualitativa del fenomeno (Sezione 2), ho condotto lo studio sul corpus *ItWac* (Sezione 4). La domanda alla quale ho provato a rispondere è la seguente: è possibile interrogare un corpus POS-taginato con alcune *query* in modo tale da riuscire a estrarre dal corpus le varie tipologie di 'che' polivalente?

L'analisi quantitativa è stata compiuta lavorando su due livelli in parallelo: quello delle *query* e quello di strutture semantico-sintattiche cui sono riconducibili i tipi di polivalenza del 'che' individuati nella Sezione 2. Tali strutture sono servite a individuare il contesto


semantico-sintattico specifico all'interno del quale i vari tipi di 'che' polivalente si vengono a trovare. Sulla base di queste, sono state formulate le *query* con le quali si è interrogato il corpus.

Terminata l'analisi in *ItWac*, ho tratto alcune conclusioni (Sezione 5) e notato come tra le varie tipologie di polivalenza del 'che', solo quelle che impiegano 'che' di tipo giustappositivo (e dunque non di tipo restrittivo) sono rintracciabili con una buona dose di probabilità in un corpus POS-taginato; i 'che' polivalenti di tipo restrittivo non sembrano invece 'ingabbiabili' solo su base contestuale in un corpus POS-taginato.

Obiettivo di questo lavoro è stato aprire una discussione in merito alla possibilità di indagare fenomeni (socio)linguisticamente significativi, o tratti emergenti dell'italiano contemporaneo, in maniera quantitativa. Con tale fine, ho cercato di capire se e in quale modo il 'che' polivalente possa emergere da un corpus taginato per *part of speech*. Evidentemente, però, valutare limiti e possibilità dell'estrazione di un fenomeno linguisticamente così sfuggente da una risorsa linguistica non è che l'inizio del lavoro. Passo successivo, una volta trovato il modo di individuare in maniera affidabile il 'che' polivalente facendo riferimento solo al suo contesto linguistico, sarà quello di porsi il concreto problema dell'annotazione (automatica) di tale *che*. In *ItWac*, seguendo il *tagset* di Baroni et al.⁶³, per il 'che' esiste un *tag* a livello di *part of speech* specifico ([tag="CHE"]). Si è visto nella Sezione 5 come sia stata evidenziata una certa probabilità di individuazione statistica su base contestuale dei casi di polivalenza del 'che' con valore giustappositivo o di introduttore di frase. A un annotatore automatico può dunque essere data istruzione di marcare ulteriormente il [tag="CHE"] come polivalente nel caso in cui il 'che' si trovi in uno dei contesti semantico-sintattici efficaci individuati nella Sezione 4⁶⁴.

Più complesso è il discorso per i 'che' restrittivi, non individuabili con una buona base statistica tramite *query*. Come notato, il problema è che generalmente il 'che' è referente di un elemento della frase in sostituzione di un pronome relativo obliquo e che questa referenza non è individuabile in un corpus unicamente POS-taginato. Unica possibile soluzione, nel caso di una risorsa del genere, è quindi la marcatura di questa referenza tramite annotazione manuale (50).

(50) L'amico che stavo parlando un attimo fa è una vecchia conoscenza.


[tag="CHE:poli"]

Un'ultima precisazione che può essere rilevante è di natura terminologica. Vista l'importante distinzione riscontrata tra 'che' giustappositivi e restrittivi, è lecito chiedersi se essa

⁶³ M. Baroni – S. Bernardini – A. Ferraresi – E. Zanchetta, *The WaCky Wide Web*.

⁶⁴ Questo genere di marcatura, secondo il *tagset* di *ItWac*, in analogia con quella in uso per le forme verbali (es. [tag="VER:fin"]) potrebbe presentarsi come [tag="CHE:poli"], in modo tale da permettere una ricerca per generico 'che' tramite l'operatore booleano *([tag="CHE.*"]).

debba o meno essere resa a livello formale nell'etichetta del *tag*⁶⁵. Ad ogni modo, data la natura già di per sé sfuggente del 'che' polivalente, un'eccessiva specificazione potrebbe risultare controproducente aumentando il rischio di errori (anche di *tagging*). Non da ultimo, sono stati menzionati casi – come il (13) – i quali non appartengono definitivamente all'una piuttosto che all'altra categoria.

Ringraziamenti

Sono profondamente grato a Caterina Mauri (Università di Bologna) per gli attenti consigli e i preziosi spunti di riflessione che mi ha offerto durante la stesura di questo lavoro.

⁶⁵ Sempre in analogia con quanto in uso per le forme verbali ([tag="VER:fin:cli"], forma finita del verbo con clitico), essa potrebbe essere [tag="CHE:poli:gapp"] in un caso, [tag="CHE:poli:res"] nell'altro.

FACOLTÀ DI SCIENZE LINGUISTICHE E LETTERATURE STRANIERE
L'ANALISI LINGUISTICA E LETTERARIA

ANNO XXV - 1/2017

EDUCatt - Ente per il Diritto allo Studio Universitario dell'Università Cattolica
Largo Gemelli 1, 20123 Milano - tel. 02.72342235 - fax 02.80.53.215
e-mail: editoriale.dsu@educatt.it (produzione)
librario.dsu@educatt.it (distribuzione)
redazione.all@unicatt.it (Redazione della Rivista)
web: www.analisilinguisticaeletteraria.eu

ISSN 1122 - 1917



9 788893 352093