

Valentina Bambini

Trattamenti formali e computazionali del Topic nel/del Discorso.

Rassegna critica

(lavoro in fase di sviluppo)

Il presente contributo costituisce un'esplorazione nell'ambito dei trattamenti formali proposti per la nozione di *Discourse Topic*, elaborati da un lato in seno ad un recente filone di ricerca, definito talvolta *Formal Pragmatics* talaltra *Discourse Semantics*, interessato all'interfaccia tra pragmatica e semantica, e dall'altro lato in seno al settore della linguistica computazionale che si occupa di *Discourse Modelling*. Il punto di partenza di una tale indagine è da ricondursi alla necessità di disporre di una definizione precisa, non intuitiva (ma neppure contro-intuitiva) della nozione di "Topic del Discorso", che sembra essere il perno del flusso di informazioni nella conversazione spontanea. Nella prima sezione preciserò i risultati da me raggiunti in precedenza studiando la Struttura Informazionale dell'enunciato, risultati dai quali la presente ricerca muove, rappresentando un'estensione dell'ambito di indagine (dall'enunciato al Discorso). In particolare, l'indagine precedentemente condotta ha dimostrato che nel parlato spontaneo prevalgono sensibilmente gli enunciati *topicless*, costituiti dal solo Comment, rispetto a quelli provvisti di Topic. Partendo da questo dato, ci si chiede se il Comment di un enunciato *topicless* sia del tutto svincolato da una qualche entità linguistica esprimente la sua *Aboutness*. L'ipotesi che appare maggiormente plausibile è che un tale Comment sia da rapportarsi ad un Topic di livello superiore all'enunciato, ovvero ad un Topic del Discorso, nozione molto chiara all'intuizione di qualsiasi parlante, ma assai sfuggente per il linguista interessato alla modellizzazione della sua portata semantico-pragmatica. In sede di introduzione sarà inoltre precisato lo statuto delle due sotto-discipline a cui mi sono rivolta per la ricerca sul *Discourse Topic*, ossia la Pragmatica Formale e l'Analisi Computazionale del Discorso. Saranno poi descritte alcune tra le principali proposte interpretative per il Topic del Discorso, in particolare tre modelli formali di stampo prettamente teorico dovuti rispettivamente a van Kuppevelt (1991), Roberts (1996) e Asher (1993) (sezione 2), un modello formale orientato computazionalmente, sebbene non implementato, quale quello proposto da Grosz & Sidner (1986) (sezione 3), ed infine l'implementazione vera e propria elaborata da Hearst (1997) (sezione 4). Confronterò poi i lavori trattati, evidenziando come tuttora manchi una formalizzazione efficace, di qualsivoglia orientamento, per il trattamento empirico dei meccanismi (interattivi) di costruzione semantico-pragmatica del Topic del Discorso, ma segnalando quali spunti teorici possano tuttavia essere tratti dalla letteratura esistente (sezione 5). A conclusione della rassegna, sarà fornita una tabella

comparativa che classifica i modelli presentati secondo i parametri nozionali ritenuti più rilevanti.*

1. Introduzione

Nel titolare questo lavoro, non ho saputo scegliere tra le espressioni “Topic *del* Discorso” e “Topic *nel* Discorso”.¹ Se da un lato era chiaro ciò che io aspiravo a trovare, ossia una caratterizzazione formale del Topic *del* Discorso, dall’altro lato non è precisamente il Topic del Discorso ciò che la mia indagine ha riscontrato. Infatti, molto spesso si è trattato di una nozione di Topic di frase inserito in un modello del Discorso, dunque di un Topic *nel* Discorso, in altre parole di una “contestualizzazione” del Topic di frase.² Tornerò a chiarire la presenza della doppia preposizione nel titolo al termine di questa introduzione, dopo aver descritto brevemente il punto di partenza della mia indagine sul Topic del Discorso e gli ambiti verso i quali mi sono rivolta.

1.1. Il punto di partenza: il Modello di Struttura Informazionale Estesa

In Bambini (2001) ho indagato come si presenta strutturato il contenuto informativo degli enunciati parlati, proponendo un modello della cosiddetta *Information Structure* che per molta parte rielabora la “Teoria della Lingua in Atto” presentata in Cresti (2000). Partendo da due assunti fondamentali quali (i) ogni Enunciato veicola un’ilocuzione e (ii) esiste un isomorfismo tra unità tonali e unità informative, ovvero ad ogni unità tonale corrisponde un’unità informativa, il modello si articola intorno ai seguenti cardini: (iii) l’ilocuzione compiuta dall’Enunciato è veicolata dall’unità tonale/informativa di Comment; (iv) l’ambito di applicazione dell’ilocuzione veicolata dal Comment è espresso dall’unità tonale/informativa di Topic, che gode della proprietà di esprimere l’*Aboutness* del Comment. Inoltre, il modello elaborato in Bambini (2001) tenta di specificare che tipo di legame sussista tra Topic e Comment all’interno dell’enunciato, affermando che (v) Topic e Comment sono legati da una relazione di Predicazione semantico-pragmatica riconducibile al principio di Pertinenza (*Relevance*), inteso come operante non tra l’enunciato ed il contesto, come nell’originaria formulazione griceana, ma, ricorsivamente, all’interno dell’enunciato tra Topic e Comment.³ Un tale modello, dove la Struttura

* Parte del materiale di questo lavoro è stato presentato durante uno dei “Seminari del venerdì pomeriggio”, anche detti “Seminari dei perfezionandi”, tenuto presso la Scuola Normale Superiore di Pisa nel dicembre 2002. La mia ricerca sul *Discourse Topic* è attualmente in corso, e non sono ancora giunta all’elaborazione di un modello soddisfacente per il computo della costruzione semantico-pragmatica del Topic del Discorso. Pertanto, il presente lavoro costituisce una presentazione dei trattamenti del Topic del Discorso reperibili nella letteratura esistente, corredata di bilancio critico e di spunti di ricerca. Nella trattazione è escluso il contributo di Berthoud (1996), segnalatomi da Massimo Moneglia a stesura quasi ultimata del presente articolo.

¹ Nel presente lavoro, verrà utilizzata l’iniziale maiuscola sia per il termine “Topic” sia per il termine “Discorso”. Nel primo caso, la scelta della maiuscola si ricollega ad una tradizione di studi funzionalista, rappresentata già in Halliday, che indica con la maiuscola il carattere di funzioni delle unità informative; nel secondo caso, la maiuscola serve a caratterizzare il valore tecnico del termine Discorso, in contrapposizione al suo uso nella lingua comune.

² L’espressione “contestualizzazione del Topic di frase” mi è stata suggerita dal professor Pier Marco Bertinetto nel corso della discussione seminariale di questo lavoro.

³ Si noti come la relazione di Pertinenza intra-enunciato, se considerata dal punto di vista del Topic, viene a coincidere con la ben nota proprietà del Topic di esprimere l’*Aboutness*, ossia di essere *about of*

Informazionale ingloba nozioni tratte dalla teoria degli atti linguistici e dalla pragmatica di impostazione griceana, è stato definito *Expanded Information Structure Model*, “Modello di Struttura Informazionale Estesa”, dove l’estensione è da intendersi come ampliamento del bagaglio teorico, non limitato alle nozioni tradizionalmente coinvolte nello studio della Struttura Informazionale.⁴

Il modello suddetto è stato testato su un campione di parlato spontaneo consistente in una conversazione informale tra più partecipanti (titolato “La cena delle zucchine”, tratto dal corpus LABLITA esemplificato in Cresti 2000, vol. II). I 674 enunciati del campione sono stati annotati informazionalmente in unità tonali/informazionali di Topic e di Comment.⁵ L’analisi ha mostrato che l’81% degli enunciati è privo di Topic (547 privi di Topic contro 87 con Topic). Osservando un dato tanto saliente, si può dedurre che nel parlato spontaneo prevalgono gli enunciati senza Topic e si possono formulare le seguenti assunzioni: (vi) il Comment gode di priorità informazionale a livello di enunciato, ovvero a livello di enunciato ricopre una funzione informazionalmente più essenziale rispetto a quella del Topic; infatti, non si danno enunciati privi di Comment, dal momento che esso veicola l’illocuzione, necessaria per l’esistenza stessa dell’enunciato; (vii) il Topic svolge una funzione informazionale che, perlomeno nell’ambito dell’enunciato, è opzionale.

Riflettendo sui punti (vi) e (vii), nasce una domanda cruciale: gli enunciati senza Topic sono costituiti da Comment del tutto privi di Aboutness, cioè non coinvolti in alcuna relazione di Pertinenza che leghi l’illocuzione ad una Aboutness? Poiché la Pertinenza, che sembra ricoprire una funzione strutturante nel Discorso e, ricorsivamente, nell’enunciato, è una nozione intrinsecamente relazionale e binaria, pare implausibile che i Comment degli enunciati privi di Topic non abbiano alcuna controparte, ossia alcun campo di applicazione. L’idea con cui si concludeva il lavoro presentato in Bambini (2001) è che negli enunciati cosiddetti *topicless* (o “tetrici”, in altre tradizioni) la Pertinenza sussista comunque, ma proceda dal Comment dell’enunciato ad un livello di strutturazione linguistica superiore a quello dell’enunciato, vale a dire al livello del Discorso. In altre parole, vi sono configurazioni informazionali in cui un Comment di Enunciato si lega attraverso una relazione di Pertinenza ad un Topic del Discorso. Per cui, oltre alla Pertinenza intra-enunciato, che lega Topic e Comment all’interno dell’enunciato, viene riconosciuta una Pertinenza intra-discorsiva che lega il Comment di enunciato al Topic del Discorso, agendo da ponte tra la micro-struttura di enunciato e la macro-struttura del Discorso.

Si osservi l’esempio seguente (sempre da “La cena delle zucchine”, Cresti 2000, vol. II, 281), i cui enunciati sono tutti *topicless*, eppure è intuitivamente molto chiaro quale sia il campo di applicazione, ovvero l>Aboutness delle illocuzioni espresse dai vari Comment.⁶

⁴ La dicitura “Struttura Informazionale” viene qui preferita a “Struttura Informativa”, dal momento che ritengo utile distinguere, teoricamente e terminologicamente, tra ciò che è la partizione dell’enunciato in Topic e Comment (Struttura Informazionale) e ciò che invece riguarda lo statuto di datità vs novità (con)testuale dei costituenti dell’enunciato (statuto informativo). Si tenga però presente che nella letteratura in italiano generalmente si parla di “Struttura Informativa”, spesso non distinguendo la modulazione in Topic e Comment dal grado di datità contestuale dei costituenti (tipicamente, si afferma che Topic=datato e Comment=nuovo).

⁵ In realtà, il modello in Bambini (2001), conformemente a Cresti (2000), prevede, oltre a Topic e Comment, anche le unità tonali/informazionali di Appendice e di Ausilio Dialogico, non direttamente rilevanti per gli scopi del presente articolo.

⁶ Gli enunciati sono trascritti in una versione implementata del formato CHAT (*Code for the Human Analysis of Transcript*), sviluppata in Cresti (2000) per permettere, oltre all’annotazione prosodica, anche quella illocutiva e informazionale. I simboli “/” e “//” indicano la fine di un’unità tonale/informazionale rispettivamente non terminale e terminale. La chiusura di un’unità terminale (ossia la fine di un’enunciato) può essere indicata anche

- (1)
1. *VER: hai fatto un corso di acquerelli? ma / stai ancora facendo / o ...
 %add: rivolto a KAT, appena tornata dal corso di acquerelli
 %ill: Richiesta di conferma // Richiesta di informazione
 %inf: Comment // Ausilio Dialogico / Comment scandito in due unità tonali
 2. *KAT: no // ho finito l'ultima lezione stasera //
 %ill: Risposta negativa // Spiegazione
 %inf: Comment // Comment
 3. *DAN: ce l'hai portati a vedere?
 %ill: Domanda totale
 %inf: Comment
 4. *KAT: sì / ha' voglia
 %ill: Risposta positiva
 %inf: Ausilio Dialogico / Comment
 5. *VER: Mingro / mangia le tue!
 %add: rivolgendosi a MIC (che si chiama Mingrone di cognome) che le stava rubando le fragole
 %ill: Rimprovero
 %inf: Ausilio Dialogico / Comment
 6. *KAT: però si guardan di là / eh?
 %com: KAT si riferisce alla confusione sul tavolo da pranzo, per cui gli acquerelli potrebbero sporcarsi
 %ill: Avvertimento
 %inf: Comment / Fatico

Nonostante non vi sia alcun Topic di enunciato, nessuno dei partecipanti alla conversazione ha avuto alcuna difficoltà nel capire quale ambito di applicazione avesse ciascun enunciato; in altre parole, a tutti era ben chiaro “di che cosa si stesse parlando” in ogni stadio della conversazione. È intuitivamente evidente che ogni enunciato (con l'eccezione di 5, che costituisce la ripresa temporanea di un argomento precedentemente oggetto di conversazione, ossia “le fragole”) si applica ad una complessa entità semantico-pragmatica, stabilita in 1 e successivamente negoziata dai partecipanti, che possiamo definire Topic del Discorso. È possibile compiere una facile operazione interpretativa ed enucleare come si struttura semantico-pragmaticamente il Topic del Discorso nell'esempio precedente. Nel momento in cui viene proferito l'enunciato 1, si stabilisce un ambito di applicazione riferito agli “acquerelli”, che viene articolato dai partecipanti in “il corso di acquerelli”, “guardare gli acquerelli”, “sporcare gli acquerelli”, ecc.

L'operazione interpretativa compiuta sopra gode di notevole plausibilità, e tale plausibilità poggia sull'intuizione derivante dalla competenza (pragmatica) di parlanti. Tuttavia, il linguista che voglia studiare il Topic del Discorso deve modellare l'intuizione, individuando i meccanismi di costruzione del significato semantico-pragmatico del Topic del Discorso. Va precisato che la nozione di Topic del Discorso si caratterizza come un

da “?” e “...”. Le righe %add e %comm riportano informazioni contestuali per la comprensione dell'enunciato; la riga %ill presenta l'annotazione illocutiva, basata su un repertorio di circa 80 illocuzioni, motivato prosodicamente e raggruppabile in cinque tipi principali (Rifiuto, Asserzione, Direzione, Espressione, Rito); la riga %inf mostra l'annotazione informazionale in Topic, Comment, Appendici e Ausili Dialogici. Si tenga presente che mentre sulla riga %inf possono apparire più unità tonali/informazionali per enunciato, la riga %ill riporta sempre una sola illocuzione per enunciato, relativa all'unità tonale/informazionale di Comment.

complesso nucleo di significato, dunque un'entità altamente semantica; tuttavia si tratta per molta parte di *pragmatic meaning*, ossia di significato stabilito contestualmente e interattivamente. Pertanto, nell'interpretazione semantico-pragmatica del Topic del Discorso è necessario rendere conto del fatto che esso si rinnova ad ogni stadio della conversazione a seconda di fattori (con)testuali, legati soprattutto all'interazione tra i parlanti. Inoltre, bisogna determinare quale materiale linguistico esplicitamente enunciato in precedenza contribuisca alla costruzione del Topic del Discorso, e soprattutto quale materiale *non* esplicitamente enunciato, ma reperibile dal contesto linguistico e situazionale, confluisca nell'agglomerato semantico-pragmatico del Topic del Discorso. Infatti, la fondamentale differenza tra il Topic di enunciato e il Topic del Discorso risiede, oltre che nel riferirsi a due unità linguistiche di estensioni differenti, nel fatto che il primo è sempre esplicitamente espresso, mentre il secondo può comporsi di materiale non esplicito. Ad esempio, possiamo intuitivamente dire che l'enunciato 6 si riferisce al sotto-Topic "sporcare gli acquerelli", la cui determinazione è possibile soltanto su base contestuale, sapendo che si era ad una cena e la tavola era sporca.

Prima di procedere nell'indagine sui tentativi di modellizzazione del Topic del Discorso, ne fornirò una definizione intuitiva, che tuttavia fa emergere la sua complessità semantico-pragmatica. Per evitare problemi di denotazione semantica in senso tradizionale, distinguiamo il Topic del Discorso dall>Aboutness del Discorso, intendendo con la seconda l'agglomerato semantico-pragmatico che costituisce l'oggetto di conversazione e definendo il primo come la realizzazione linguistica (non necessariamente esplicita in ogni sua componente) della seconda.

- (2) ABOUTNESS DEL DISCORSO : proprietà ideativa del Discorso, consistente in un conglomerato semantico-pragmatico di referenti, stati, eventi che circoscrivono l'ambito di applicazione del complesso degli enunciati, e si determina rinnovandosi in ogni stadio del Discorso sulla base di ciascuno degli enunciati (e rispettivi contesti di enunciazione) precedentemente generati e negoziati nella dinamica complessiva del processo comunicativo.
- (3) TOPIC DEL DISCORSO (definizione provvisoria): realizzazione linguistico-cognitiva dell>Aboutness del Discorso che si protrae lungo la dinamica del Discorso distribuendosi concretamente in parti di enunciato (non necessariamente in unità topicali) o in enunciati interi, e/o creando schemi cognitivi non espressi esplicitamente.

1.2. Il Topic del Discorso nell'Analisi del Discorso

Assumendo come stella polare l'ideale di una definizione precisa del Topic del Discorso, mi sono rivolta inizialmente all'ambito della *Discourse Analysis* (compresa la sua variante denominata *Conversation Analysis*), all'interno del quale il *Discourse Topic* fa storicamente la sua comparsa. Spesso, nei lavori di Analisi del Discorso, viene menzionata la nozione di Discourse Topic, definito come *what the discourse is about*, intendendo talvolta ciò di cui parla il Discorso nel suo complesso (qualcosa di analogo al titolo di un testo), talaltra ciò di cui parla una data porzione di Discorso.⁷ Dopo un esame complessivo

⁷ Si vedano Keenan & Schieffelin (1976), Kintsch & van Dijk (1978), Beaugrande & Dressler (1981), tra i primi studi a menzionare il Discourse Topic. Si tenga presente che spesso nella letteratura si parla di *Text Topic*, inteso come sinonimo di Discourse Topic. L'oscillazione tra le due espressioni è dovuta al fatto che alcuni studiosi parlano di "Testo" (soprattutto nella tradizione europea della *TextLinguistik*) altri di "Discorso" (soprattutto nella tradizione americana della *Discourse Analysis*). In questa sede verrà preferita la dicitura "Topic del Discorso" a

della letteratura, ho potuto constatare una generale vaghezza delle definizioni, e soprattutto la debolezza dei metodi di identificazione del Topic del Discorso, affidati quasi del tutto all'intuizione. Paradigmatica al riguardo è la seguente affermazione contenuta in Brown & Yule (1983), un testo classico nell'Analisi del Discorso.

- (4) The notion of "topic" is clearly an intuitively satisfactory way of describing the unifying principle which makes one stretch of discourse "about" something and the next stretch "about" something else, for it is appealed to very frequently in the discourse analysis literature [...] Yet the basis for the identification of "topic" is rarely made explicit. (Brown & Yule 1983, 69-70)

Per citare soltanto un altro studio paradigmatico sull'argomento, menzionerò il lavoro di Chafe (1994), dove il Discourse Topic è indagato dettagliatamente nel suo snodarsi parallelamente al "flusso di coscienza" e nella sua negoziazione da parte dei partecipanti, basandosi su di una definizione senz'altro perspicace, ma ad ogni modo intuitiva.

- (5) One of the things that seem intuitively true of conversations is that they focus on different topics, in this sense, at different times, moving from one topic to another. We can think of each topic as an aggregate of coherently related events, states, and referents that are held together in some form in the speaker's semiactive consciousness. (Chafe 1994: 121)

Di fatto, molta parte della ricerca in Analisi del Discorso si è concentrata sullo studio dei meccanismi di negoziazione e di sotto-articolazione del Topic del Discorso, tralasciando il problema del suo computo semantico-pragmatico. In particolare, sono state sviluppate varie strategie per l'individuazione empirica dei Topic del Discorso che si susseguono nel testo. A questo riguardo, molta attenzione è stata rivolta alla fase del *Topic shift*, ossia il momento di chiusura di un Topic e passaggio al Topic successivo.

- (6) It has been suggested [...] that instead of undertaking the difficult task of attempting to define "what a Topic is", we should concentrate on describing what we recognize as a Topic shift. That is, between two contiguous pieces of discourse which are intuitively considered to have two different "topics", there should be a point at which the shift from one Topic to the next is marked. If we can characterize this marking of topic-shift, then we shall have found a structural basis for dividing up stretches of discourse into a series of smaller units, each on a separate Topic [...] The burden of analysis is consequently transferred to identifying the formal markers of topic-shift in discourse. (Brown & Yule 1983: 94-95)

Senza dubbio l'Analisi del Discorso ha il merito di aver indagato le proprietà del Topic del Discorso, conformemente peraltro alla vocazione descrittiva di questa linea di ricerca. In particolare, mettendo insieme i vari contributi, è possibile stilare un elenco delle proprietà del Topic del Discorso : (a) la globalità, ossia il fatto che alla sua costruzione partecipa l'intero complesso degli enunciati del Discorso (*any utterance beyond backchannel items contributes to a conversational topic*, Lenk 1988: 25); (b) la concentrabilità, ossia il fatto che esso può apparire espresso in maniera concentrata in singole parti di enunciato, non necessariamente comprese in unità informative di Topic

quella di "Topic del Testo", dal momento che l'unità linguistica presa in considerazione è il Discorso, inteso come prodotto dinamico, piuttosto che il Testo, il quale sembra più appropriato in riferimento al prodotto del Discorso, inteso come finito e statico.

(*encapsulating Discourse Topic*, Downing 1997); (c) la gerarchizzabilità in più sotto-Topic del Discorso (*macro/global Topic vs micro/local Topic*, Tracy 1984; *superordinate Topic*, van Oosten 1985); (d) l'interattività/negoziabilità nel corso della conversazione (*navigation by schema*, Chafe 1994); (e) l'identificabilità, ossia il fatto che i parlanti sono in grado di identificarlo (fondatezza psicolinguistica a tale identificabilità è fornita in Giora *et al.* 1996).

Inoltre, sono state ben individuate le fasi del *Topic Management*, che possono essere schematizzate nelle seguenti tappe, da intendersi in maniera non lineare: (a) la fissazione (*Topic establishment*); (b) lo sviluppo, che può articolarsi in mantenimento, elaborazione, (ri)focalizzazione, ripresa, ecc.; (c) lo scioglimento; (d) il cambio (*Topic shift*), che può essere solamente temporaneo, in altre parole una digressione, o definitivo; (e) la sotto-articolazione, ossia la creazione di sotto-Topic (caratteristica già rappresentata al punto c dell'elenco precedente come "gerarchizzabilità").

L'Analisi del Discorso, sebbene fornisca utili descrizioni sulle proprietà pragmatico-linguistiche e strutturali del Topic del Discorso, non offre alcun modello per la sua definizione semantico-pragmatica. Da qui, il mio interesse per settori della linguistica più "formali", alla ricerca di un procedimento rigoroso, un algoritmo magari, per descrivere il Topic del Discorso.

1.3. I trattamenti formali del Topic

È necessario precisare che cosa io intenda in questa sede per "formalizzazioni" del Topic del Discorso.⁸ Non ricerco trattamenti in termini strettamente logico-matematici: personalmente, ritengo che le formalizzazioni estreme mal si applichino ad un prodotto del linguaggio naturale, tanto più se si considera la variante del Discorso non pianificato. Piuttosto, essendo il Topic del Discorso una nozione altamente semantica, seppur di significato non (soltanto) vero-condizionale bensì pragmatico, ritengo che sia possibile trattarlo con i metodi formali già applicati negli studi di semantica. Dunque, aspiro ad una teoria formalizzabile nel senso che dovrebbe essere: (i) innanzitutto fondata su basi non intuitive; (ii) poi traducibile in regole (o tendenze) espresse con una notazione limpida, magari affine a quella utilizzata per la logica e la matematica; (iii) infine, eventualmente, implementabile in uno o più algoritmi.

Gli ambiti in cui il Topic del Discorso è stato trattato in termini formali sono essenzialmente due: il primo, composito al suo interno, si colloca all'interfaccia tra pragmatica semantica, il secondo afferisce alla linguistica computazionale e si occupa della modellizzazione del Discorso.⁹ In sede introduttiva, è necessario spendere qualche parola per chiarire l'origine e lo statuto epistemologico di queste due linee di ricerca, entrambe piuttosto recenti e ancora in fase di sviluppo.

⁸ Si confronti la posizione di van Kuppevelt, uno degli autori qui trattati. Egli precisa di non intendere l'espressione *formal characterization/definition* nel senso matematico o logico, bensì in riferimento a *the essential, though not per se operational, characteristics of (the definiendum or) that which is characterized* (nota 5 in van Kuppevelt 1995: 111).

⁹ Naturalmente, trattamenti formali del Topic sono stati di recente proposti anche in studi sintattico-semantici di orientamento generativista (cfr. Erteshik-Shir 1997 e Büring 1997). Essi non verranno considerati in questa sede, in quanto non direttamente concernenti la dimensione del Discorso e peraltro inapplicabili agli enunciati parlati.

1.3.1. L'interfaccia (formale) tra pragmatica e semantica

A partire dai primi anni Novanta, emergono problemi legati a residue incertezze nella definizione dell'ambito della linguistica pragmatica: sembra che la distinzione classica tra pragmatica e semantica, secondo cui la prima si occuperebbe degli aspetti del significato non considerati dalla semantica intesa come vero-condizionalità, sia eccessivamente netta e spesso fuorviante. I pragmaticisti si accorgono che nei lavori di pragmalinguistica resta penalizzata la dimensione semantica e cominciano a fare uso di strumenti propri della semantica, soprattutto di tipo logico-formale. In questo quadro, nasce la cosiddetta *Formal Pragmatics*.

Allo stesso tempo, da parte dei semanticisti, ci si rende conto della necessità di considerare aspetti contestuali-discorsivi nel computo del significato (Peregrin 1999 parla di *pragmaticization of semantics*). Comincia a delinearsi così un ambito di studi, definito *Discourse Semantics* (e talvolta *Dynamic Semantics*), che indaga i fenomeni semantici connessi alla dimensione del Discorso /contesto.

Sebbene sia possibile distinguere tra lavori orientati più pragmaticamente e lavori orientati più semanticamente, *Formal Pragmatics* e *Discourse Semantics* si occupano sostanzialmente dei medesimi oggetti linguistici, ovvero dei fenomeni che si collocano alla cosiddetta "interfaccia" tra semantica e pragmatica, intendendo la prima come il dominio formale del significato e la seconda come il dominio del Discorso -contesto. Ci si occupa sia di fenomeni semantici reperibili al micro-livello di enunciato ma interpretabili soltanto (con)testualmente, come la presupposizione, le implicature, la referenza anaforica, sia del macro-livello del Discorso considerato nel suo complesso valore semantico. In particolare, nascono vari modelli dinamici per la modellizzazione della costruzione (semantica) del Discorso.¹⁰

Anche nell'ambito degli studi sull'organizzazione dell'informazione (*Information Structure*), tradizionalmente collocati sotto l'ombrello della pragmatica, si assiste, su scala ridotta, ad un ampliamento di prospettiva in direzione semantica. Già negli anni Ottanta, erano state affrontate questioni al limite tra semantica e pragmatica, come il rapporto tra presupposizione e Focus e la referenza anaforica. Tuttavia la prospettiva era eminentemente frastica. Negli anni Novanta invece, la dimensione del Discorso diventa imprescindibile e ci si concentra soprattutto sul ruolo discorsivo dell'articolazione della frase in Topic-Comment. In precedenza, gli studi di pragmatica sulla Struttura Informazionale consideravano lo snodarsi di Topic e Comment secondo parametri di datità vs novità dei costituenti. Ora, invece, nel quadro della Pragmatica Formale, si afferma sempre di più il cosiddetto approccio *questioning*, che poggia su una versione ampliata al Discorso delle teorie semantiche tradizionalmente assunte per la domanda. Il Topic rappresenta una domanda e gode dell'interpretazione semantica classica delle domande (si veda Hamblin 1973), ossia denota l'insieme (semantico) delle possibili risposte che possono essere date alla domanda in considerazione. Tuttavia, tale insieme è determinato contestualmente, sulla base del Discorso precedente e della situazione contestuale. Si noti che l'approccio *questioning* non ha nulla a che fare con il tradizionale criterio empirico che consiste nell'usare la domanda "di che cosa si parla?" per la determinazione del Topic. Nei modelli

¹⁰ Per avere un'idea dei fenomeni oggetto di studio in Pragmatica Formale/Semantica del Discorso, si veda Turner (1999), che fa parte di una recente collana dedicata all'interfaccia semantica-pragmatica. Si veda anche il volume di Kadmon (2001), intitolato per l'appunto *Formal Pragmatics. Semantics, Pragmatics, Presupposition and Focus*.

di Pragmatica Formale / Semantica del Discorso le domande strutturano il Discorso, ne fanno parte e non sono poste *a posteriori* dall'analista a fini interpretativi.¹¹

1.3.2. Il Discourse Modelling computazionale

Il secondo ambito in cui il Topic ha subito trattamenti formali è quello della linguistica computazionale. Sempre a partire dagli anni Novanta, anche sul versante computazione si possono constatare gli effetti dovuti all'ormai raggiunto sviluppo della pragmatica. Il *Natural Language Processing*, dopo essersi occupato molto di sintassi e di lessico, affronta il *parsing* discorsivo. Nel 1995 Frank van Eynde afferma che "the future of Natural Language Processing is in the processing of discourse" (introduzione a van Eynde & Allegranza 1995: 7). In pratica, si può dire che lo studio del Discorso trova uno sviluppo computazionale nel cosiddetto *Discourse Modelling*.¹²

All'interno dei modelli computazionali, un'importante distinzione deve essere fatta tra le proposte *theory-driven* e le proposte *implementation-driven*. Da un lato vi sono computazionalisti rivolti alla creazione di modelli "ingegneristici" della struttura del Discorso, nelle cui proposte l'implementabilità è assunta soltanto in linea teorica, senza essere verificata empiricamente; dall'altro lato vi sono computazionalisti impegnati nell'implementazione vera e propria, ossia nella creazione di sistemi informatici per il trattamento e la generazione di discorsi/dialoghi. Generalmente, i modelli computazionali di entrambe le tipologie si ispirano a teorie proposte in seno alla Discourse Analysis tradizionale, alla Pragmatica Formale e alla Semantica del Discorso. Chiaramente, i modelli *theory-driven* hanno un'architettura piuttosto complessa e articolata, non priva di sottigliezze formali, e la loro portata teorica è generalmente non inferiore a quella dei modelli del Discorso proposti in Analisi del Discorso e pragmatica Formale. Per contro, i modelli implementati sono decisamente più snelli e semplificano in maniera netta rispetto alle teorie prese a riferimento.

Nell'ambito del Discourse Modelling, compresa la sua variante di *Dialogue Modelling*, si fa largo anche lo studio del Topic del Discorso. Per quanto riguarda le proposte di stampo teorico, emerge la necessità di tener presente la strutturazione dell'informazionale (o flusso di informazioni) nella modellizzazione del Discorso. Per quanto riguarda invece i trattamenti empirici e implementati del Discorso, ci si accorge che il recupero delle informazioni contenute in una base di dati, oppure reperibili da un sistema di dialogo, è guidato da non meglio precisate idea di Topic (si tratta dello studio dell'*Information Retrieval*). Inoltre, si fa strada l'idea che la segmentazione dei testi debba avvenire secondo unità caratterizzate dal medesimo Topic (è l'ambito cosiddetto *TextTiling*).

Già da questo breve paragrafo, si può notare che l'aspetto semantico del Topic del Discorso non rappresenta una preoccupazione per i computazionalisti, i quali si

¹¹ In realtà, l'approccio questioning vanta lontane origini. Si vedano Vennemann (1975) e Bartsch (1976), dove il Topic è identificato con una delle presupposizioni definite dalla domanda. In Klein & Von Steutterheim (1987) il Topic è inteso come l'alternativa, la scelta tra due o più possibilità come risposta alla domanda posta. Dunque, fin dalle origini sembra che l'approccio questioning e un'idea di *Alternative Semantics* si intreccino con il Topic. In questa sede esamineremo l'approccio questioning come appare in van Kuppevelt (1991 e segg.) e in Roberts (1996 e segg.), ma devono essere menzionate anche le proposte di von Fintel (1994) e Büring (1997).

¹² A questo proposito, si veda l'introduzione di Walker & Moore (1997) al numero speciale di *Computational Linguistics* dedicato a *Empirical Studies in Discourse*. Si veda anche il cap. 6, intitolato *Discourse and Dialogue*, in Varile & Zampolli (1997).

concentrano maggiormente sui suoi aspetti pragmatico-ingegneristici, ossia sul ruolo che il Topic del Discorso svolge nella strutturazione del Discorso stesso.

1.4. Bilancio preventivo

In sede introduttiva, ritengo utile chiarire immediatamente che né la Pragmatica Formale né la linguistica computazionale hanno finora fornito un modello in grado di monitorare in maniera precisa e formalizzata, i meccanismi di progressiva costruzione del valore semantico-pragmatico del Topic del Discorso. Nei modelli elaborati in seno alla Formal Pragmatics/Discourse Semantics lo studio del Topic è inserito in una più ampia teoria del Discorso, generalmente piuttosto articolata e complessa. Le definizioni proposte per la nozione di Topic sono raffinate dal punto di vista formale-semantico, grazie alla trovata del questioning. Si fa uso di formalismi per rendere conto della semantica del Topic, ma ciò che viene rappresentato formalmente è un Topic di frase, più che un Topic del Discorso. Gli approcci computazionali con orientamento teorico prestano maggiore attenzione alla dimensione del Discorso, proponendo modelli ingegneristici delle dinamiche discorsive. Tuttavia, il significato (pragmatico) del Topic del Discorso, ossia la determinazione della sua portata semantica, non rappresenta un centro di interesse nella costruzione di tali modelli. La perifericità degli aspetti semantici è ancora più evidente nelle implementazioni vere e proprie, dove il significato del Topic del Discorso sembra essere ridotto ad un *pattern* lessicale.

Alla luce di quanto detto sopra, concluderò precisando meglio il senso della doppia preposizione del titolo. Se nella Discourse Analysis/Conversation Analysis avevo riscontrato la vaghezza delle definizioni e la debolezza criteri di identificazione del Topic del Discorso, nei vari trattamenti formali ho trovato una certa (con) fusione di piani. Dal momento che la frase è immersa nel Discorso/contesto, spesso la distinzione tra Topic di frase e Topic di Discorso va sfumandosi, a vantaggio però del Topic di frase. In altre parole, si mette a fuoco un Topic di frase contestualizzato nel Discorso. Certamente, la nozione di Topic di frase si caratterizza per un'intrinseca biplasia, essendo allo stesso tempo un costituente di frase ed un referente del Discorso; tuttavia tale biplasia non deve offuscare la nozione di Topic del Discorso, che gode di autonomia dal livello della frase/enunciato.

Procederemo ora valutando alcune dei principali modelli proposti nei settori disciplinari sopra introdotti. Chiaramente, tra tutti i modelli disponibili sono stati selezionati quelli in cui il Topic del/nel Discorso ricopre un ruolo degno di nota. Cominceremo con la Pragmatica Formale / Semantica del Discorso, dunque considerando approcci formali con orientamento teorico, rappresentati dalle proposte di van Kuppevelt (1991 e segg.), Roberts (1996 e segg.) e Asher (1993). Passeremo poi agli approcci formali con orientamento computazionale, distinguendo tra le proposte teoriche e le proposte implementate. Per illustrare un modello computazionale a sfondo teorico sarà descritto il modello del Discorso di Grosz & Sidner (1986); per esemplificare una vera e propria implementazione considereremo l'algoritmo a base lessicale di Hearst (1997).

2. Approcci formali con orientamento teorico

2.1. Il modello di Jan van Kuppevelt

Una delle prime formalizzazioni del Topic è stata proposta da Jan van Kuppevelt in vari lavori pubblicati a partire dai primi anni Novanta.¹³ La proposta di van Kuppevelt assume una relazione diretta tra struttura del Discorso, Topic e domande: il Discorso deriva la sua coerenza strutturale precipuamente dalla struttura Topic-Comment, interna al Discorso stesso e organizzata gerarchicamente, che è il risultato di un procedimento di induzione contestuale di domande *topic-forming* esplicite o implicite. Tale procedimento induttivo rappresenta una componente essenziale del processo di produzione del Discorso.

In tale quadro, viene proposta una nozione unitaria del termine Topic, che comprende sia la nozione di *sentence Topic* sia quella di Topic di unità del Discorso superiori (ma il *Discourse Topic* è a parte, come vedremo).

- (7) The notion [of topic] presupposes that a discourse unit U, a sentence or a larger part of a discourse, has the property of being, in some sense, directed at a selected set of discourse entities (a set of persons, objects, places, times, reasons, consequences, actions, events or some other sets) and not diffusely at all discourse entities that are introduced or implied by U. This selected set of *entities in focus of attention* is what U is about and is called the Topic of U. In the framework of this formal characterization the complementary notion of Comment is characterized as that which is newly asserted of the Topic of U. (van Kuppevelt 1995a: 112, corsivo mio)¹⁴

Partendo da una tale definizione, viene elaborata una teoria che si differenzia dalle comuni ricerche riferite specificamente al livello di frase oppure al livello di Discorso.

Oltre alla definizione in (7), serve una caratterizzazione operativa del Topic, un criterio per distinguere le entità Topic dalle altre entità del Discorso. Nella letteratura sono stati usati svariati criteri al riguardo, talvolta basati sulla sintassi (il Topic si identifica con il soggetto di frase), basati sullo statuto informativo dei sintagmi (il Topic corrisponde al materiale contestualmente dato), ecc. A questo riguardo, van Kuppevelt adotta l'innovativo criterio del *questioning*, ossia il criterio della domanda. La suddetta nozione uniforme di Topic si qualifica, oltre che come *context-dependent*, come *question-based*, ovvero si assume che il Topic sia il risultato di una domanda e che venga chiuso quando (il parlante assume che) alla domanda corrispondente sia stata fornita una risposta soddisfacente.

- (8) Every contextually induced explicit or implicit (sub)question Qp that is answered in discourse constitutes a (sub)topic Tp. Tp is that which is being questioned; a set of singular or plural (possibly non-existent)¹⁵ discourse entities (or a set of ordered *n*-tuples of such

¹³ Il modello viene sviluppato per esteso in van Kuppevelt (1991). I successivi lavori si concentrano ciascuno su aspetti più specifici. Qui farò riferimento soprattutto a van Kuppevelt (1995a). Molto interessante, ma non sviluppato in questa sede è van Kuppevelt (1996), dove la questione, molto dibattuta in pragmatica, sulle implicature scalari (derivanti dalla prima massima griceana della Quantità) viene trattata riconducendola ad inferenze legate al Topic.

¹⁴ Si noti l'espressione "entities in Focus" riferita al Topic, che richiama il *Focus of attention* delle scienze cognitive e della linguistica computazionale. Infatti, nella sezione 3, vedremo che il modello, teorico ma orientato computazionalmente, proposto da Grosz & Sidner (1986) parla di *Focus space* in riferimento a ciò che noi chiamiamo Topic del Discorso.

¹⁵ Tale caso si verifica, ad esempio, quando l'entità di Topic è espressa da espressioni del tipo *nobody nothing*.

entities in the case of a n -fold question) from which one is selected as an answer to Q_p . Comment C_p is provided by this answer and names or specifies the entity asked for. (van Kuppevelt 1995a: 113)

Tra tutte le entità che vengono introdotte o sono implicate nel Discorso, soltanto quel sottoinsieme di entità oggetto esplicito o implicito di una domanda possono essere caratterizzate come Topic.

Dal punto di vista semantico, il Topic T_p , ovvero *that which is being questioned*, si caratterizza come l'intensione del termine Topic (che si dà nell'analisi sintattica della domanda), ossia come l'insieme delle possibili estensioni di tale termine. Per contro, il comment C_p , *that which is asked for*, è l'estensione del termine Topic nella situazione attuale. Al momento della domanda, l'estensione del termine Topic nella situazione attuale $T_p(S_{act})$ è ancora *un(der)determined*, implicando che il Topic range $\rho'(T_p)$ di cui $T_p(S_{act})$ è un elemento non contiene ancora un valore unico.

$$(9) \quad T_p(S_{act}) \in \rho'(T_p) \quad | \rho'(T_p) > 1 |$$

La riduzione dell'indeterminatezza del Topic si ottiene con la risposta alla domanda Q_p , ovvero con il Comment. Si osservi il seguente esempio, in cui, allo stadio della domanda, il range di T_1 contiene le tutte le possibili entità del Discorso (Alan e Brian), per poi assumere un valore preciso nel momento della risposta (Alan).

$$(10) \quad \begin{array}{l} Q_1: \quad \text{Who is (the one who) is laughing?} \\ \quad T_1(S_{act}) = -? \\ \quad \text{Assumendo che qualcuno stia ridendo ed assumendo un dominio contestuale } D = \\ \quad \{ \text{Alan, Brian} \}, \\ \quad T_1 = \{ \langle S_1, \{ \text{Alan} \} \rangle, \langle S_2, \{ \text{Brian} \} \rangle, \langle S_3, \{ \text{Alan, Brian} \} \rangle \} \\ \quad \text{Parlando in termini di range,} \\ \quad T_1(S_{act}) \in \rho'(T_1) \\ \quad \rho'(T_1) = \{ \{ \text{Alan} \}, \{ \text{Brian} \}, \{ \text{Alan, Brian} \} \} \\ \\ A_1: \quad \text{Alan is laughing.} \\ \quad T_1(S_{act}) = C_1 \\ \quad T_1(S_{act}) = \{ \{ \text{Alan} \} \} \\ \quad | \rho'(T_1) = 1 | \end{array}$$

Naturalmente, nella realtà della produzione linguistica, i discorsi raramente si articolano secondo esplicite coppie domanda-risposta. Spesso, le domande che danno origine ai Topic, soprattutto nel caso dei testi monologici non viene formulata alcuna domanda. Tuttavia van Kuppevelt ritiene che sia sempre possibile ricostruire le domande implicite nel testo, secondo un processo di induzione contestuale.¹⁶ Dunque, i principi in gioco nella produzione del Discorso restano invariati nel caso delle domande implicite.

Le entità del Discorso che rappresentano i Topic non sono sciolte, bensì organizzate gerarchicamente. Il Discorso infatti si sviluppa secondo un processo di *questioning* la cui gerarchia coinvolge principalmente tre parametri: *feeder*, *topic-constituting questions* e

¹⁶ In van Kuppevelt (1991) viene fornito un algoritmo per la ricostruzione delle domande implicite, basato su fattori strettamente contestuali, di cui però lo stesso van Kuppevelt riconosce i limiti.

subtopic-constituting subquestions. Per *feeder* van Kuppevelt intende un evento, linguistico o non linguistico, che dà origine ad una *main topic-constituting question*. Se linguistico, si tratta di un'unità priva di Topic oppure il cui Topic non è più prominente al momento della domanda. Se non linguistico, si tratta di un pensiero o di un evento che occorre nella situazione linguistica e che è percepito per lo meno da uno dei partecipanti (si pensi ad un rumore che dà origine ad una domanda/Topic "Che cosa è stato?"). Le domande originate dai *feeder* sono dotate di uno statuto di autonomia, ovvero non sono dipendenti da altre domande, e possono essere sia implicite sia esplicite. L'esempio seguente mostra un *feeder* F_1 che origina una *main-topic constituting question*.

- (11) F_1 : A well-know subsidy book publisher is searching for manuscripts.
 Q_1 : What kind of manuscripts?
 $T_1(S_{act}) \in \{X \mid X \subseteq \text{TYPES_OF_MANUSCRIPTS}\}$
 A_1 : Fiction and non-fiction.
 $T_1(S_{act}) = \{\text{Fiction, Non-fiction}\}!$ (van Kuppevelt 1996: 399)¹⁷

Vi sono poi domande non autonome, gerarchicamente subordinate. Si tratta delle *subtopic constituting questions*, che si generano da una risposta insoddisfacente ad una precedente domanda, insoddisfacente o dal punto di vista qualitativo o dal punto di vista quantitativo.

- (12) An explicit or implicit subquestion Q_p is contextually induced as the result of a quantitatively or qualitatively unsatisfactory answer A_{p-n} given to a preceding higher-order question Q_{p-n} . In the former case the unsatisfactory answer results in a quantitative underdeterminedness of the actual Topic extension $T_{p-n}(S_{act})$, implying that the comment value provided by A_{p-n} is incomplete. In the latter case the underdeterminedness of $T_{p-n}(S_{act})$ is of a qualitative nature, e.g. because A_{p-n} is not specific enough. (van Kuppevelt 1996: 400)

Segue un esempio con una sottodomanda Q_2 che sorge da una risposta A_1 insufficiente dal punto di vista quantitativo.

- (13) F_1 : A well-known subsidy book publisher is searching for manuscripts.
 Q_1 : What kind of manuscripts?
 $T_1(S_{act}) \in \{X \mid X \subseteq \text{TYPES_OF_MANUSCRIPTS}\}$
 A_1 : Fiction and non-fiction will be considered.
 $T_1(S_{act}) \in \{T_1(S_i) \in \rho'(T_1) \mid \{\text{Fiction, Non-fiction}\} \subseteq T_1(S_i)\}$
 Q_2 : What else?
 $T_2(S_{act}) \in \{X \mid X \subseteq (\text{TYPES_OF_MANUSCRIPTS} - \{\text{Fiction, Non-fiction}\}) \wedge |X| = 1\}$
 A_2 : Poetry and travel.
 $T_2(S_{act}) = \{\text{Poetry, Travel}\}!$
 $T_1(S_{act}) = \{\text{Fiction, Non-fiction, Poetry, Travel}\}!$ (van Kuppevelt 1996: 400)

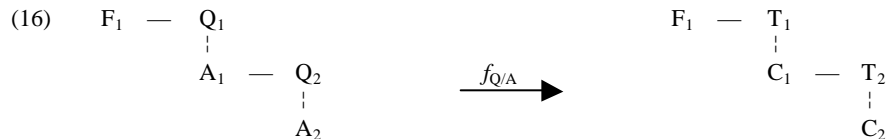
Il procedimento di questioning che struttura il Discorso è regolato da due principi, il primo riguardante l'ordine in cui le sottodomande sono contestualmente indotte, il secondo concernente il processo di esaurimento del Topic. van Kuppevelt li qualifica entrambi come

¹⁷ Con il punto esclamativo viene indicata la chiusura del Topic.

principi di coerenza, nel senso di Sperber & Wilson, ovvero di coerenza ottimale (in termini di effetti contestuali) rispetto al minimo sforzo di processing.

- (14) *Principle of Recency*: In a well-formed discourse every explicit or implicit subquestion Q_p is asked as the result of an answer A_{p-n} , which is the most recent unsatisfactory answer to a preceding question Q_{p-n} .
- (15) *Dynamic Principle of Topic Termination*: If an explicit or implicit (sub)question Q_p is answered satisfactorily, the questioning process associated with it comes to an end. As a consequence, Topic $T_p (T_p = f_{Q/A}(Q_p))$ loses its actuality in discourse.

La struttura dinamica del Discorso definita con le nozioni sopra descritte può essere “mappata”, ossia rappresentata nel suo snodarsi utilizzando la generale funzione $f_{Q/A}$ che connette il dominio domande-risposte con il dominio Topic-Comment. Per l'esempio (13), la mappatura risulta essere la seguente.



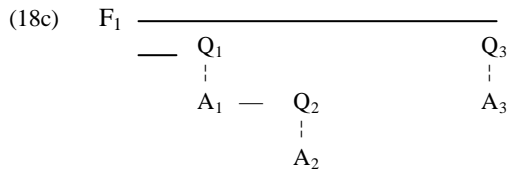
La teoria di van Kuppevelt non distingue a priori tra Topic di frase e Topic del Discorso, bensì assume una nozione uniforme di topic, ammettendo però Topic di ordine più alto (rappresentati da domande) e Topic di ordine più basso (rappresentati da sotto-domande). Esiste poi un Topic di ordine ancora più alto: il *Discourse Topic*, inteso qui come l'insieme dei Topic originati dallo stesso feeder, quindi un insieme di domande appartenenti alla categoria delle *main topic-constituting questions*.

- (17) A Discourse Topic DT_i is defined by the set of all topics T_p that are constituted as the result of one and the same feeder F_i ($DT_i = \{T_p \mid T_p \text{ constituted a.r.o. } F_i\}$). As such DT_i is a set of main, higher order topics usually hierarchically comprising lower topics.

Si dà anche la possibilità in cui da un feeder nasca un solo Topic, che, in tale evenienza, coincide con il Discourse Topic. È il caso dell'esempio (13). Segue ora un esempio di Discourse Topic più complesso, composto da due domande di ordine principale, originate entrambe dallo stesso feeder.

- (18a) F_1 : Yesterday a jury of investigation came to the conclusion that the 31 casualties of the fire in the King's Cross London underground station died as the result of an accident and not as the result of negligence.
 Q_1 : How did people react to the outcome of the investigation?
 A_1 : Relatives of the victims rejected it.
 Q_2 : Why?
 A_2 : They are of the opinion that the jury did not do their job well.
 Q_3 : What is the consequence of this outcome?
 A_3 : The consequence is that the further prosecution of the officials of London Regional Transport is ruled out.

$$\begin{aligned}
 (18b) \quad DT_1 &= \{T_p \mid T_p \text{ constituted a.r.o. } F_1\} = \\
 &= \{T_1, T_3\} \\
 &(T_1 = f_{Q/A}(Q_1) \text{ and } T_3 = f_{Q/A}(Q_3))
 \end{aligned}$$



2.1.1. Riflessioni in itinere

Il modello elaborato da van Kuppevelt si sviluppa ulteriormente, considerando aspetti sui quali non mi soffermerò in questa sede. Vorrei però puntualizzare alcune caratteristiche specifiche di questo modello, che esulano dalla valutazione comparativa complessiva che fornirò nella sezione finale.

Innanzitutto, mi pare degna di nota l'idea di adottare una prospettiva uniforme nel definire la nozione di Topic, senza distinguere aprioristicamente tra Topic di frase e Topic del Discorso. Di fatto, però, viene a configurarsi un modello in cui i Topic sono legati alla frase (anche se intesa come collocata nella più ampia dimensione contestuale), mentre i Discourse Topic sono definiti in maniera meccanicamente compositiva, come somma di Topic frastici. Una teoria davvero uniforme del Topic dovrebbe essere più duttile, e contemplare i casi di Topic di frase in cui è concentrato un Topic del Discorso e i casi di Topic del Discorso che non si danno meccanicamente.

Il *feeder* è una trovata interessante, che meriterebbe maggiore studio. In particolare, è buona l'idea che ci siano come dei "precursori" del topic, ma non sono sicura che essi siano totalmente *topicless* e non debbano affatto comparire nel computo del valore semantico del Topic stesso. Mi pare piuttosto che dovrebbero essere considerati come facenti parte del Discourse Topic.

Vorrei poi segnalare la somiglianza tra la struttura Q/A di van Kuppevelt e il modello di Struttura Tematica elaborato da Daneš negli anni Settanta nell'ambito della seconda Scuola di Praga. Daneš studia i Temi di frase in rapporto al testo di cui le frasi fanno parte, soffermandosi sulla loro concatenazione e gerarchia. La struttura del testo viene vista come una sequenza di temi che si snodano secondo meccanismi di progressione differenti (lineare, a salti, ecc.). Chiaramente, nella prospettiva di Daneš, la dimensione semantica non è considerata affatto.

Segnalo infine la stretta connessione esistente tra la coppia di *Principle of Recency* e *Dynamic Principle of Topic Termination* da un lato e le massime griceane dall'altro. A prima vista sembra che i due principi siano legati soprattutto alle Massime di Quantità e Qualità, dato che le sotto-domande si generano da risposte insoddisfacenti quantitativamente o qualitativamente.¹⁸ Tuttavia, a ben guardare, si tratta di principi di coerenza del testo, che governano la *Relevance* tra le domande e le risposte.

¹⁸ Infatti, van Kuppevelt (1996) risolve, nel quadro della sua teoria, il problema delle implicature scalari derivanti proprio dalla prima sotto-Massima di Quantità.

2.2. Il modello di Craige Roberts

Secondo Roberts (1996, 1998a), prima di studiare i primitivi della cosiddetta *Information Structure*, necessitiamo di una teoria pragmatica unificata che definisca in maniera rigorosa le nozioni rilevanti del Discorso, in maniera tale che esse possano regolare il flusso di informazioni nel Discorso stesso. Solo con alle spalle una simile teoria è possibile studiare i fenomeni connessi con il Topic ed il Focus della frase. A questo scopo, Roberts procede alla costruzione di una teoria generale della struttura del Discorso.

Molto schematicamente, le idee che sottostanno alla teoria del Discorso sono: (i) il Discorso come *language game*, strutturato su relazioni domanda/risposta (idea che fa capo ai lavori di Wittgenstein, Lewis, Carlson); (ii) il linguaggio come intenzionale, che riflette *goals* e *plans* degli interlocutori (sulla scorta dei lavori di Grice e dalla *Planning Theory* sviluppata in Intelligenza Artificiale); (iii) la conversazione come avente per scopo primario la *communal inquiry*, ossia la condivisione di informazioni: gli interlocutori mirano a estendere il *common ground* per includere più informazioni. Quest'ultima idea risale ai lavori di Stalnaker, dove il *common ground* è caratterizzato come l'insieme di proposizioni che gli interlocutori assumono o credono sia vero, essendo ogni proposizione un insieme di mondi (quelli in cui la proposizione è vera), mentre l'intersezione dei *common ground* degli interlocutori dà il *context set*. Dalla congiunzione delle tre idee base sopra menzionate deriva una concezione del Discorso riassumibile nella frase *language is a game of intentional inquiry* (Roberts 1998a: 114). In altre parole, il Discorso è organizzato intorno ad una serie di mete e piani conversazionali che i partecipanti, intenzionalmente, sviluppano al fine di raggiungere uno scambio informativo.

Gli elementi del gioco del Discorso identificati da Roberts sono:

(a) *goals*, miranti a raggiungere un accordo sul modo in cui le cose sono nel mondo, ovvero a massimizzare il *common ground* degli interlocutori, riducendo il *context set* ad un singolo insieme, il mondo attuale;

(b) regole, ossia limitazioni sul comportamento linguistico degli interlocutori. Sono di due tipi: convenzionali, ovvero propriamente linguistiche (sintattiche, composizionali, ecc.), e conversazionali, ovvero non linguistiche ma che seguono da considerazioni razionali in vista degli scopi del gioco (le massime griceane);

(c) mosse, vale a dire comportamenti linguistici che rispettano le regole e sono classificati in accordo alle loro relazioni agli scopi del gioco; ne esistono due tipi: mosse *set-up* (domande, le quali fissano gli scopi immediati del Discorso, ossia rispondere alle domande stesse) e mosse *pay-off* (asserzioni, le risposte alle domande, risposte che conseguono gli scopi del Discorso stabiliti). Ogni mossa ha sia un contenuto presupposto sia un contenuto proferito (per una domanda, un insieme di alternative; per un'asserzione, ciò che è asserito). Una mossa può essere accettata o respinta.¹⁹

(d) strategie, ossia modi di porre in sequenza mosse, in vista delle loro relazioni logiche, per raggiungere scopi accettati. Consistono nella creazione di sotto-domande.

Un ruolo cruciale nel sistema è ricoperto dalle mosse *set-up*, ossia dalle domande. Seguendo Hamblin (1973), viene assunta la semantica standard per la domanda, secondo cui essa denota l'insieme di proposizioni che possono costituirne la risposta, insieme

¹⁹ Si noti che le mosse non sono atti linguistici, bensì gli oggetti semantici usati negli atti linguistici. Un atto linguistico è l'atto di proferire una mossa.

chiamato da Roberts *Q-alt(ernative) set*.²⁰ La denotazione di una domanda equivale al suo *Q-alt set*, che è l'insieme delle proposizioni che possono essere derivate astraendo dal *wh-phrase*, e applicando questa astrazione a tutte le cose che appartengono al modello del Discorso, sono dello stesso tipo e sono denotate dal sintagma *wh-*.

- (19) Interpretation of a question “? α ”:
 $| ?\alpha | = Q\text{-alt}(\alpha)$
- (20) The Q-alternatives corresponding to utterance of a clause a:
 $Q\text{-alt}(a) = \{p: \exists u^{i-1}, \dots, u^{i-n} \in D [p = |\beta| (u^{i-1}) \dots (u^{i-n})]\}$, where
 α has the logical form $wh_{i-1}, \dots, wh_{i-n}(\beta)$, with $\{wh_{i-1}, \dots, wh_{i-n}\}$ the (possibly empty) set of *wh*-elements in α ,
and D is the domain of the model for language, suitable sortally restricted (e.g. to humans for *who*, non-humans for *what*).²¹

L'esempio seguente mostra l'interpretazione della domanda *Who did Mary invite?*, che, in un dominio del Discorso D che contempla soltanto *Mary*, *Alice* e *Grace*, consiste nelle insieme di due alternative: *Mary invited Alice* e *Mary invited Grace* (Roberts 1996: 98).

- (21) Who did Mary invite?
logical form ?Who (?x. Mary invited x)
interpretation |?(Who did Mary invite)|
 $|?(Who did Mary invite)| = Q\text{-alt} (Who (?x.Mary invited x)) =$
 $= \{p: \exists u \in D [p = |\lambda x.Mary\ invited\ x| (u)]\} =$
 $= \{Mary\ invited\ u : u \in D\}$
 $D = \{Mary, Alice, Grace\}$
 $Q\text{-alt} = \{Mary\ invited\ Alice, Mary\ invited\ Grace\}$

La novità sta nel contributo che le domande apportano alla Struttura Informativa. In sintonia con i lavori di Carlson, Roberts sostiene che le domande stabiliscano le mete che guidano la *inquiry* dell'interlocutore. Se una mossa *set-up* viene accettata dagli interlocutori, questo li incarica di una meta comune, quella di trovare la risposta, scegliendo tra le possibili alternative. La domanda accettata diventa l'immediata *Question Under Discussion* (QUD), che non è altro che il Topic.

- (22) If a *set-up* moves, or question, is accepted by the interlocutors, this commits them to a common ground, finding the answer, choosing among the proffered alternatives. The accepted question becomes the immediate Topic of discussion, which I also call the **question under discussion** (QUD). (Roberts 1998: 115)

In pratica, il Topic è una mossa *set-up* che stabilisce una meta del Discorso e deve essere soddisfatta. In altre parole è una domanda, da intendersi però non come atto

²⁰ Più precisamente, una domanda denota l'insieme di proposizioni ottenute sostituendo tutti i valori del giusto tipo alle espressioni *wh-* che si danno nella traduzione logica della domanda. Per esempio, *Who arrived?* si traduce logicamente come $?[arrived (who)]$.

²¹ La formulazione si legge pressoché nel seguente modo: il *Q-alt* della domanda α che ha la forma logica *wh-* (β) è dato dall'insieme delle proposizioni p tali che esiste un certo numero di entità u appartenenti al dominio del Discorso D e le proposizioni p sono composte da (β) u .

linguistico, bensì come oggetto semantico che denota un insieme di proposizioni possibili a cui bisogna rispondere.

All'interno del modello del Discorso elaborato, la Struttura Informazionale viene intesa semplicemente come il flusso dell'informazione che costituisce la meta del Discorso stesso. Ne deriva che è concepita come svincolata da limiti di dimensioni, quindi non legata alla frase, bensì estesa nel Discorso.

- (23) The conception I will develop here is one in which information structure is literally a structure on information – on the inquiry pursued in discourse and the information which that inquiry yields, and not on the utterances or sequences of utterances used to present it.

La Struttura Informazionale evolve dinamicamente, ma è possibile immaginarla come statica, in un dato momento, e definirla formalmente come una tupla composta da mosse (domande e risposte), una relazione di precedenza $<$, il *common ground* degli interlocutori, sulla base del quale viene stabilito il dominio del Discorso D , e le QUD (Roberts 1996: 99-100).

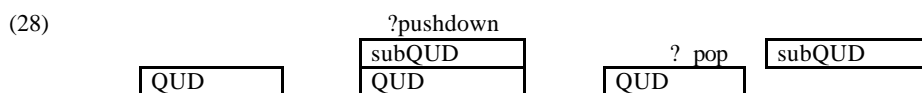
- (24) The information structure for a discourse D is a tuple,
InfoStr_D = $\langle M, Q, A, <, Acc, Cg, QUD \rangle$ where:
M is the set of (setup and payoff) moves in the discourse.
Q \subseteq **M** is the set of questions (setup moves) in **M**, where a question is a set of propositions.
A \subseteq **M** is the set of assertions (payoff moves) in **M**, where an assertion is a set of possible worlds.
 $<$ is the precedence relation, a total order on **M**; $M_i < m_k$ iff M_i is made/uttered before m_k in D ; the order of any two elements under $<$ will be reflected in the natural order on their indices, where for all $M_i, i \in \mathbb{N}$.
Acc \subseteq **M**, is the set of accepted moves.
CG is a function from **M** to sets of propositions yielding for each $m \in \mathbf{M}$ the common ground of D just prior to the utterance of m . Further, we require that:
a. for all $m_k \in \mathbf{M}$, $CG(m_k) \supseteq \cup_{i < k} (CG(M_i))$,
b. for all $m_k \in \mathbf{M}$, $CG(m_k) \supseteq \{M_i : i < k \text{ and } M_i \in \text{Acc} \setminus Q\}$, and
c. for all $m_k, M_i \in \mathbf{M}$, $i < k$,
i. the proposition that $M_i \in \mathbf{M}$ is in $CG(m_k)$,
ii. if $M_i \in Q$, the proposition that $M_i \in Q$ is in $CG(m_k)$,
iii. if $M_i \in A$, the proposition that $M_i \in A$ is in $CG(m_k)$,
iv. if $M_i \in \text{Acc}$, the proposition that $M_i \in \text{Acc}$ is in $CG(m_k)$,
v. for all propositions $p \in CG(m_i)$, the proposition that $p \in CG(m_i)$ is in $CG(m_k)$, and
vi. whatever the value of $QUD(M_i)$, the proposition that that is the value of $QUD(M_i)$ is in $CG(m_k)$.
QUD, the question-under-discussion stack, is a function from **M** (the moves in the discourse) to ordered subset of $Q \cap \text{Acc}$ such that for all $m \in \mathbf{M}$:
a. for all $q \in Q \cap \text{Acc}$, $q \in QUD(m)$ iff
i. $q < m$ (i.e., neither m or any subsequent questions are included), and
ii. $CG(m)$ fails to entail an answer to q and q has not been determined to be practically unanswerable.
b. $QUD(m)$ is (totally) ordered by $<$.
c. for all $q, q' \in QUD(m)$, if $q < q'$, then the complete answer to q' contextually entails a partial answer to q .

Nella Struttura Informazionale agisce un principio di *Relevance*, inteso come una regola conversazionale, preposta alla gestione delle mosse, che prescrive che ogni enunciato (ogni mossa) contribuisca allo scopo in corso, ossia al fornire una risposta alla QUD. Sono considerate mosse pertinenti sia le risposte alle domande/Topic sia le sotto-domande/sub-Topic che possono generarsi all'interno di una strategia per rispondere alla domanda principale.

- (25) A **partial answer** to a question q is a proposition which contextually entails the evaluation – either true or false – of at least one element of $Q\text{-alt}(q)$. A **complete answer** is a proposition which contextually entails an evaluation for each element of $Q\text{-alt}(q)$.
- (26) A move m **addresses** a question q iff m either introduces a partial answer to q (m is an assertion) or is part of a strategy to answer q (m is a question).
- (27) A move will be **Relevant** in the Information Structure of the discourse in which it occurs iff it addresses the question under discussion at the time it is proffered.

L'insieme delle QUD viene modellato usando un *push-down store*, chiamato *QUD stack*. Ricordiamo che una pila è una struttura per il riconoscimento di linguaggi a memoria stratificata. La pila nasce per il trattamento di linguaggi di tipo 2, il cui riconoscimento richiede una struttura dotata di memoria. Essa ha fondamentalmente due operazioni: *push* e *pop*, la prima delle quali rappresenta l'essenza della pila, denominata per l'appunto *push-down stack*, *push-down automata*. Il funzionamento della pila prevede che, dato un certo stato 1, l'analisi prosegue "spingendo" (fase di *push*) in cima alla pila un nuovo stato 2, senza però eliminare l'1. Lo stato 2, una volta trattato ed esaurito, viene eliminato (fase di *pop*) restituendo lo stato 1.

Roberts utilizza il meccanismo della pila per il riconoscimento non di linguaggi di tipo 2 ma di testi dialogici, e assumendo che gli stati siano rappresentati da QUD. Potremmo dunque rappresentare il meccanismo a pila della Struttura Informazionale nel modo seguente:



Per ogni data mossa m , la funzione QUD comporta l'insieme ordinato di tutte le domande, non ancora risolte ma risolvibili, accettate in Q nel momento di enunciazione di m . Quando accettiamo una domanda, la aggiungiamo sulla cima della pila (*to the top of the stack*). La sua relazione con ogni domanda precedentemente sulla cima della pila è garantita dalla regola della *Relevance*. Quando viene data la risposta alla domanda, essa viene rimossa dalla pila (*popped from the stack*) e la risposta viene aggiunta al common ground. Se si assiste alla creazione di sotto-domande per risolvere una domanda iniziale, la pila funziona finché l'ultima sotto-domanda viene risolta, fornendo la risposta alla domanda iniziale. A quel punto il QUD stack resta vuoto.

Per mostrare come lavora il QUD stack, viene fornito un esempio molto semplificato, che presuppone un modello del Discorso D_0 con due soli individui, *Hilary* e *Robert*, e due soli tipi di cibo, *bagels* e *tofu*.

- (29) 1. Who ate what?
 a. What did Hilary eat?
 a_i. Did Hilary eat bagels?
 Ans(a_i) Yes.
 a_{ii}. Did Hilary eat tofu?
 Ans(a_{ii}) Yes.
 b. What did Robin eat?
 b_i. Did Robin eat bagels?
 Ans(b_i) No.
 b_{ii}. Did Robin eat tofu?
 Ans(b_{ii}) Yes.

La funzione QUD per l'esempio sopra si presenta nel modo seguente:²²

- (30) QUD(1) = \emptyset
 QUD(a) = $\langle 1 \rangle$
 QUD(a_i) = $\langle 1, a \rangle$
 QUD(Ans(a_i)) = $\langle 1, a, a_i \rangle$
 QUD(a_{ii}) = $\langle 1, a \rangle$
 QUD(Ans(a_{ii})) = $\langle 1, a, a_{ii} \rangle$
 QUD(b) = $\langle 1 \rangle$
 QUD(b_i) = $\langle 1, b \rangle$
 QUD(Ans(b_i)) = $\langle 1, b, b_i \rangle$
 QUD(b_{ii}) = $\langle 1, b \rangle$
 QUD(Ans(b_{ii})) = $\langle 1, b, b_{ii} \rangle$

Il modello di Roberts, a differenza di quello di van Kuppevelt, ha già trovato un'implementazione, realizzata da Davis (2000). Il modello è stato sfruttato per la costruzione di un sistema per il trattamento di dialoghi *query based* e *task-oriented*, in cui il partecipante umano chiede alla macchina informazioni per realizzare una prenotazione alberghiera. L'architettura del sistema è di tipo modulare; un modulo è rappresentato dal QUD stack di Roberts. In più vi sono vari algoritmi più specifici per la risoluzione di operatori particolari (un pronome, una descrizione definita, una presupposizione). Si tratta di un algoritmo estremamente pregevole, sebbene la sua efficacia sia rigidamente limitata alla tipologia testuale dei dialoghi *task-oriented* e *query-based*.

2.3. Accenni alla Segmented Discourse Representation Theory di Nicolas Asher

La *Segmented Discourse Representation Theory* (SDRT) proposta in Asher (1993) nasce dalla *Discourse Representation Theory* (DRT) di Kamp & Reule (1981) per risolvere problemi di referenza ad oggetti astratti.²³ Si tratta di una DRT arricchita con un insieme di

²² Roberts sviluppa anche un'analisi della nozione Focus, interpretata anch'essa secondo un approccio di *alternative semantics*. Non mi soffermerò sulla teoria del Focus, limitandomi a segnalare l'esistenza della *Question Under Discussion constraint on Focus*, secondo cui, perché la mossa risulti felice, il Focus deve costituire la risposta alla QUD.

²³ Mentre i modelli di van Kuppevelt e di Roberts sono stati delineati grosso modo nella loro totalità, della SDRT verrà qui data una visione molto parziale, trattandosi di un modello di interpretazione del Discorso assai raffinato e

discourse relations, ossia con relazioni retoriche che legano segmenti di Discorso, ispirate ai lavori di Polany, di Hobbs, e soprattutto alla *Rethorical Structure Theory* (RST) di Mann & Thompson (1986).

La SDRT interpreta la struttura del Discorso attraverso l'individuazione di *Segmented Discourse Representation Structure* (SDRS). Una SDRS consiste in un insieme di *Discourse Representation Structure* (DRS), chiamate *costituenti* della SDRS, fatte di segmenti di testo, ovvero di proposizioni, e di un insieme di condizioni su tali DRS, condizioni che asseriscono che certe relazioni del Discorso vigono tra certe DRS.²⁴ Le DRS vengono rappresentate graficamente con il sistema di *boxes* già utilizzato nella DRT; le SDRS consistono in boxes di taglia superiore inglobanti le DRS.

Studiando la concatenazione dei segmenti di Discorso, ossia i meccanismi di *updating* del Discorso, emerge anche la questione del Topic. L'*updating* prevede che ogni nuova DRS del Discorso sia in qualche modo connessa a precedenti DRS, secondo una regola di Relevance.

- (31) **Rule of Relevance:** given an SDRS a for an initial segment of a discourse D , any new constituent derived from further processing of D must be attached to some constituent of a , if D is coherent. (Asher 1993: 270)

Una delle possibili modalità di *updating* è a base topicale: la nuova DRS si attacca alla precedente istituendo una relazione topicale, ossia formulando esplicitamente il Topic di entrambe oppure elaborando il Topic della DRS di ancoraggio.²⁵ In tale quadro, il Topic è sostanzialmente un costituente, ovvero una DRS, che riassume il contenuto di altre DRS.

- (32) Topics do not contribute generally to the text's content about its subject matter. Their principal function is to contribute an important structural component of SDRSs. Topics are summary constituents for a group of constituents. A topic is a proposition that summarizes the content of a constituent in an SDRS, and it bears a particular structural relation to that constituent. (Asher 1993: 267)

La definizione in termini formali si presenta come segue (Asher 1993: 267).²⁶

ampio, che trascende la Struttura Informazionale. *A latere*, confesso una mia speranza delusa. Leggendo il volume di Asher (1993), che si propone principalmente di risolvere il problema della referenza ad oggetti astratti (proposizioni, proprietà, stati, fatti), speravo di trovare una formulazione che consentisse di interpretare i meccanismi cognitivi che sottostanno all'assunzione di tali oggetti al ruolo di referenti del Discorso, ossia di Topic. In altre parole, speravo di rinvenire un modello in grado di spiegare i meccanismi concettuali (ma anche semantico-pragmatici) che fanno sì, ad esempio, che una proposizione intera, dunque un'entità astratta, diventi un Topic del Discorso, a cui ci si può riferire anaforicamente. Invece, l'interesse di Asher si rivolge all'interpretazione degli oggetti astratti in ottica puramente semantica, e non alla questione della loro interpretazione semantico-pragmatica nel Discorso.

²⁴ Nella DRT di Kamp & Reule, invece, il Discorso si articola in DRS, dove i referenti linguistici rappresentano i costituenti e i predicati rappresentano le condizioni su tali referenti.

²⁵ Secondo Asher (1993: 276), l'*updating* topic-based può sfruttare due tipi di relazioni retoriche: *Elaboration* e *Continuation*. Nella versione della SDRT elaborata da Grabsky (2000) sfruttando apporti della *Channel Theory* e dalla *Situation Semantics*, anche la relazione retorica di *Narration* è ammessa.

²⁶ Riferisco alcune convenzioni notazionali utilizzate da Asher (1993). Le lettere con doppio contorno \mathbb{K} indicano SDRS. Le lettere greche indicano i costituenti delle SDRS, ovvero indicano DRS o subDRS. Per comodità, spesso le SDRS vengono indicate con variabili locali con doppio contorno \mathbb{K} , \mathbb{K}_a , ecc. e le DRS con variabili locali del tipo k , k_i , ecc. (Asher 1993: 263). Il simbolo β indica una relazione gerarchica di *discourse-dominance*: $a \beta \mathbb{B}$ iff a subsumes \mathbb{B} or (\mathbb{B} entails a and not a entails \mathbb{B}). In pratica, $a \beta \mathbb{B}$ quando a funziona da

- (33) Let α, β be constituent in \mathbb{K} . $\text{Topic}(\beta) = \alpha$ iff α is simple and either:
 (i) $\alpha\beta$ in \mathbb{K} (in this case α is an *explicitly represented topic*); or
 (ii) $\neg\exists\gamma < \beta$ ($\text{Continuation}(\gamma, \alpha)$ or $\text{Continuation}(\alpha, \gamma)$ or $\text{Background}(\alpha, \gamma)$) and $\forall\delta < \beta \exists R \in \text{Discourse Relation } R(\alpha, \delta)$ or $R(\delta, \alpha)$ in \mathbb{K} (in this case α is an *implicitly represented topic*).

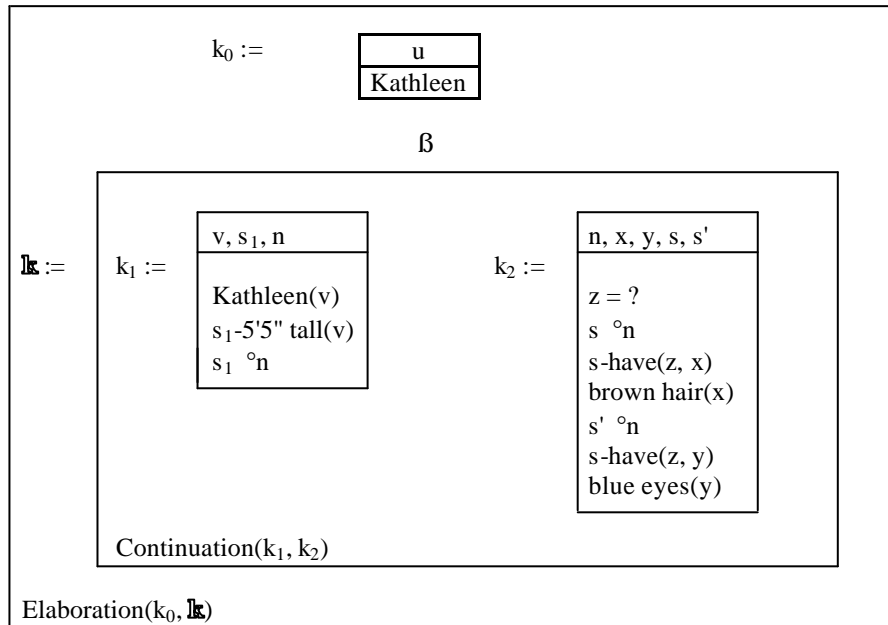
Si osservi l'esempio seguente (qui in veste semplificata, ossia ridotta a due enunciati, mentre la versione originaria Asher 1993: 276 ne prevede un terzo). Il primo ed il secondo enunciato vengono elaborati rispettivamente nella prima e nella seconda DRS, ciascuna raffigurata con un box. Tuttavia, la regola di Relevance prevede che ci sia un "aggancio" a cui ancorare il secondo enunciato. In questo caso, il secondo enunciato sfrutta l'aggiornamento a base topicale, prelevando il costituente "Kathleen" e promuovendolo allo statuto di Topic per se stessa e anche per il primo enunciato. Di conseguenza, viene costruita una SDRS che contiene le due precedenti DRS, legandole entrambe al Topic da una relazione di Elaborazione (mentre la relazione che le lega l'una all'altra è di Continuazione).

- (34) 1. Kathleen is five feet five inches tall. 2. She has brown hair and green eyes.

v, s_1, n
Kathleen(v)
s_1 -5'5" tall(v)
$s_1 \circ n$

n, x, y, s, s'
$z = ?$
$s \circ n$
s-have(z, x)
brown hair(x)
$s' \circ n$
s-have(z, y)
blue eyes(y)

riassunto o Topic statement di β . Inoltre, in riferimento all'esempio (34), si tenga presente che le lettere "s, s, ..." denotano *state discourse referents*; "n" si riferisce al momento di enunciazione e il simbolo \circ indica la contemporaneità al momento di enunciazione (per la simbologia completa, cfr. Aher 1993: 86).



Sicuramente, il costituente “Kathleen” rappresenta il Topic più plausibile nella porzione di Discorso formata dai due enunciati in (34). Tuttavia, l’esempio fa emergere uno snodo teorico poco chiaro del modello di Asher. La definizione di Topic data in (32) asserisce che si tratta di una proposizione, una proposizione riassunto, ma pur sempre una proposizione. Invece, nell’esempio il Topic è rappresentato da un referente del Discorso, un termine, non una proposizione. Può darsi che, nella sua sottile argomentazione semantica, Asher da qualche parte affermi che anche un singolo referente può costituire una DRS, dunque una proposizione; tuttavia la nozione di Topic nella SDRT è sicuramente poco chiara ed andrebbe meglio precisato a quale livello di strutturazione linguistica appartenga, se sia un termine (esplicito o implicito) od una proposizione in senso canonico.²⁷

A titolo di completezza, segnalo che in Asher & Lascarides (1998), la SDRT viene arricchita con nuove relazioni retoriche, ispirate al settore computazionale della *Plan Recognition*, al fine di renderla utilizzabile anche per i dialoghi: si tratta infatti di relazioni retoriche dialogiche e intenzionali. Un’interessante proposta di sviluppo della SDRT proviene dai lavori di Prévot, dove la SDRT estesa ai dialoghi è ulteriormente integrata con relazioni di tipo informazionale. In particolare, il Discorso si struttura in SDRS, a cui si affianca la *Topic Structure*, distinta ma strettamente connessa con la SDRS.²⁸

²⁷ Sarebbe interessante confrontare il Topic del modello di Asher con il *Link* del modello di Vallduví (1991): il primo usa la rappresentazione in boxes, il secondo usa la metafora del *file-card*, ma vi è una notevole somiglianza di fondo. Infatti, Vallduví si richiama alla *File Change Semantics* proposta da Heim all’inizio degli anni Ottanta, un modello per la rappresentazione dei referenti del Discorso che assomiglia molto alla DRT cui si richiama Asher. Un confronto tra la rappresentazione a box e la rappresentazione a file-card si trova in Hendriks (2002).

²⁸ Il rapporto tra relazioni retoriche del Discorso e Struttura Informazionale è peraltro indagato anche dallo stesso Asher in Asher & Txurruka (1995), ai quali Prévot fa riferimento.

3. Approcci formali con orientamento computazionale

I modelli “ingegneristici” della struttura del Discorso proposti in linguistica computazionale sono numerosi, senza contare che spesso i computazionalisti assumono e sviluppano modelli formulati in ambiti non specificamente computazionali (si pensi alla RST di Mann & Thompson 1986) e spesso operano fusioni di modelli (ad esempio la *Relational Discourse Analysis* (RDA) di Mooser & Moore 1996, che fonde la RST con il modello di Grosz & Sidner 1986). Verrà qui considerata soltanto la proposta, originale e innovativa, di Grosz & Sidner (1986), essendo quella in cui maggiormente la nozione di Topic del Discorso svolge un ruolo essenziale, sebbene sotto le mentite spoglie di *Focus Space*.²⁹

3.1. Il modello di Barbara Grosz & Candace Sidner

In Grosz & Sidner (1986) viene elaborata una teoria computazionale della struttura del Discorso che integra due linee di ricerca: i lavori sul *Focusing* di Barbara Grosz e i lavori sul riconoscimento di intenzioni di Candace Sidner. Si tratta di una teoria molto innovativa per l'epoca, che infatti segnò una svolta nella linguistica computazionale, sia perché sancì l'ingresso a pieno titolo dell'analisi del Discorso nel NLP sia perché introdusse due nozioni non linguistiche quali *intention* e *attention*.³⁰

Il Discorso, sia scritto sia parlato, è definito come una porzione di comportamento linguistico che coinvolge più enunciati e più partecipanti, questi ultimi distinti in *Iniziating Conversational Participant* (ICP), che produce il primo enunciato di un segmento del Discorso, e *Other Conversational Participant(s)* (OCP). La macro-struttura del Discorso si articola in tre componenti: la struttura linguistica, la struttura intenzionale e lo stato attenzionale.

(i) A livello della struttura linguistica, il Discorso si divide in segmenti. Così come le parole in una singola frase si aggregano naturalmente in sintagmi, così gli enunciati in un Discorso si aggregano naturalmente in segmenti. È possibile formulare la seguente proporzione: la parola sta al sintagma come l'enunciato sta al segmento di Discorso (chiaramente sottintendendo l'equiparazione della frase al Discorso). Un segmento è una struttura discontinua. Due enunciati consecutivi possono appartenere allo stesso segmento, ma non necessariamente. Anche due enunciati non consecutivi possono appartenere allo stesso segmento. La fondatezza della segmentazione a livello di Discorso è provata da diversi studi psicolinguistici sulla individuazione di *cues* linguistici, quali, ad esempio,

²⁹ Altri contributi da parte di computazionalisti allo studio del Topic del Discorso, senza però giungere all'elaborazione di veri e propri modelli innovativi, si trovano nel volume curato da Mitkov & Nicolov (1997): si vedano gli interventi di Kruijff & Schaake, di Nomato e di Ramm. Merita di essere segnalato anche il lavoro di Poesio & Di Eugenio (2001), dove la RDA è sviluppata per la risoluzione della referenza, venendo dunque a coinvolgere più direttamente la nozione di Topic.

³⁰ L'impatto di Grosz & Sidner (1986) fu enorme, e non solo sul NLP. Ora è diventato uno studio classico, citato in linguistica computazionale ma anche in Pragmatica Formale e Semantica del Discorso (cfr. Bos 1995, che riporta anche l'elenco delle implementazioni concretamente realizzate sul modello di Grosz & Sidner; van Benthem & ter Meulen 1997; Zuo & Zuo 2001).

pause ai confini, *speech rate* più lenta all'inizio di un segmento, cambiamenti nelle scelte tempo-aspettuali, espressioni del tipo "in primo luogo" (si tratta dei cosiddetti *discourse markers*). Così come meccanismi linguistici possono influenzare la segmentazione, anche la segmentazione può influenzare fenomeni linguistici (soprattutto di referenza). La struttura linguistica è fatta di segmenti e di *embedding relations* che sussistono tra i segmenti e che sono il riflesso delle relazioni tra gli elementi della struttura intenzionale.

(ii) La struttura intenzionale comprende intenzioni e relazioni tra di esse, e fornisce la base razionale del Discorso. Un partecipante può avere più di uno scopo nel partecipare ad un Discorso, ma si assume che uno di essi sia quello fondamentale, definito *Discourse Purpose* (DP). Ad esempio, una storia può voler intrattenere l'ascoltatore, così come descrivere un evento. Per ogni segmento del Discorso, possiamo individuare un'intenzione, definita *Discourse Segment Purpose* (DSP), che specifica come quel segmento contribuisca al conseguimento del DP generale. Ci sono molti tipi di intenzioni (stupire, impressionare). Qui si considerano quelle intenzioni che si vuole che vengano riconosciute dall'interlocutore, non quelle private. Si noti che DP e DSP condividono certe proprietà con le intenzioni di frase che Grice aveva usato per definire il significato di enunciato. Esempi di DSP sono *Intend that someone intend to perform some physical task* e *Intend that someone believe some fact*. Vi sono due fondamentali relazioni strutturali tra DSP: *dominance* (dom) e *satisfaction-precedence* (sat-prec). Si dice che DSP1 sat-prec DSP2 quando DSP1 deve essere soddisfatta prima di DSP2. Si dice che DSP2 dom DSP1 quando DSP1 contribuisce al soddisfacimento di DSP2, fornendo parte della soddisfazione di DSP2. I DSP costituiscono una lista aperta. Un modello del Discorso non può dipendere dallo scegliere le intenzioni da una lista definita. Tuttavia è essenziale per la struttura del Discorso identificare le relazioni strutturali tra le intenzioni.

(iii) Lo stato attenzionale (*attentional state*) è un'astrazione dalla nozione psicologica di *focus of attention* dei partecipanti. Il *focus of attention* registra le entità salienti, ossia le entità che sono state menzionate di recente o sono implicite nel contesto. Lo stato attenzionale astrae dal *focus of attention* dei partecipanti le entità salienti in ogni dato punto del Discorso. In altre parole, esso registra gli oggetti, le proprietà, le relazioni che sono salienti in ogni dato punto del Discorso. Tra le entità salienti si considera anche il DSP, ossia l'intenzione che caratterizza quel segmento di Discorso. I cambiamenti nello stato attenzionale dipendono dalla struttura intenzionale e dalle proprietà dell'enunciato nella struttura linguistica. Lo stato attenzionale è inerentemente dinamico³¹ ed è una proprietà del Discorso stesso, non dei partecipanti, essendo un'astrazione da una nozione psicologica. Esso è modellato da un insieme di *focus space*, ciascuno corrispondente ad un diverso stato attenzionale; vi sono poi regole di transizione che precisano le condizioni di aggiunta e cancellazione degli spazi. La collezione degli spazi disponibili in ogni momento è detta *focus structure* e il processo di modellamento è detto *focusing*. Il processo di focusing associa un *focus space* ad ogni DS; tale spazio contiene le entità salienti in quel dato DS.

Per la rappresentazione della struttura triadica del Discorso, Grosz & Sidner propongono uno *stack model*, ovvero un modello a pila.³² Nell'esempio seguente viene illustrata l'analisi della struttura di un dialogo di tipo *task oriented* e *query based*, suddividendo la stringa di testo in DS, precisando le relazioni (intenzionali) che sussistono

³¹ Non a caso il modello di Grosz & Sidner è descritto alla voce "Dynamics", curata da Muskens *et al.*, nel manuale di van Benthem & ter Meulen (1997).

³² Si veda però Walker (1996) che critica lo *stack model* e propone un *cache model*.

tra i segmenti, e raffigurando l'andamento degli enunciati con il meccanismo a pila (Grosz & Sidner 1986: 186 e segg.).

- (35a)
- | | | | |
|-----|------|----|--|
| DS1 | (1) | E: | First you have to remove the flywheel. |
| | (2) | A: | How do I remove the flywheel? |
| | (3) | E: | First, loosen the two allen head setscrews holding it to the shaft, then pull it off. |
| | (4) | A: | OK. |
| DS2 | (5) | | I can only find one screw. Where's the other one? |
| | (6) | E: | On the hub of the flywheel. |
| | (7) | A: | That's the one I found. Where's the other one? |
| | (8) | E: | About ninety degrees around the hub from the first one. |
| | (9) | A: | I don't understand. I can only find one. Oh wait, yes, I think I was on the wrong wheel. |
| | (10) | E: | Show me what you are doing. |
| | (11) | A: | I was on the wrong wheel and I can find them both now. |
| DS3 | (12) | | The tool I have is awkward. Is there another one tool that I could use instead? |
| | (13) | E: | Show me the tool you are using. |
| | (14) | A: | OK. |
| | (15) | E: | Are you sure you are using the right size key? |
| | (16) | A: | I'll try some others. |
| | (17) | | I found an angle I can get at it. |
| DS4 | (18) | | The two screws are loose, but I'm having trouble getting the wheel off. |
| DS5 | (19) | E: | Use the wheelpuller. Do you know how to use it? |
| | (20) | A: | No. |
| | (21) | E: | Do you know what it looks like? |
| | (22) | A: | Yes. |
| | (23) | E: | Show it to me please. |
| | (24) | A: | OK. |
| | (25) | E: | Good. Loosen the screw in the center and place the jaws around the hub of the wheel then tighten the screw onto the center of the shaft. The wheel should slide off. |

(35b) Primary Intentions:

- I1: (Intend E (Intend A (Remove A flywheel)))
- I2: (Intend A (Intend E (Tell E A (Location other setscrew))))
- I3: (Intend A (Intend E (Identify E A another tool)))
- I4: (Intend A (Intend E (Tell E A (How (Getoff A wheel))))))
- I5: (Intend E (Know-How-to A (Use A wheelpuller)))

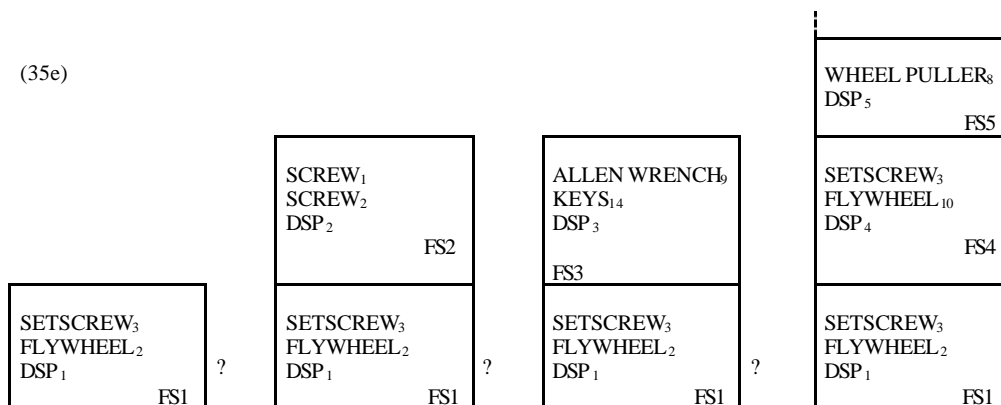
(35c) Dominance Relationships:

- I1 DOM I2
- I1 DOM I3
- I1 DOM I4
- I4 DOM I5

(35d) Satisfaction-Precedence Relationships:

- I2 SP I3
- I2 SP I4
- I3 SP I4

(35e)



3.1.1. Il Topic del Discorso nel modello di Grosz & Sidner

A questo punto, è legittimo chiedersi quale sia, nel modello di Grosz & Sidner, il ruolo del Topic del Discorso, che non è mai stato menzionato nella trattazione precedente. Grosz & Sidner (1986), confrontandosi con la letteratura esistente, accennano in al Topic del Discorso nel seguente passaggio.

(36) In the literature the concept of “topic” has appeared in many guises. [...] It appears that many of the descriptions of sentence topic correspond (though not always) to centers, while discourse topics corresponds to the DSP [Discourse Segment Purpose] of a segment or of the discourse. (Grosz & Sidner 1986: 192)

Gli autori affermano che il loro Discourse Segment Purpose corrisponde al cosiddetto Topic del Discorso. In pratica, secondo loro, il Topic di un segmento di Discorso è costituito dall’intenzione propria di quel determinato segmento di Discorso.

In realtà, aldilà della formulazione originaria, nella versione “vulgata” del modello di Grosz & Sidner, il Topic del Discorso è equiparato non tanto al singolo DSP, quanto piuttosto al focus space in cui il DSP è incluso, dunque all’insieme di oggetti, proprietà, relazioni e intenzioni registrato dal focus of attention dei partecipanti in ogni stadio della

conversazione. A titolo di esempio cito il lavoro di Zuo & Zuo (2001) sul Focus, dove, a proposito del modello di Grosz & Sidner, si afferma che il Focus della linguistica computazionale e dell'Intelligenza Artificiale corrisponde alla diffusa nozione di Topic (del Discorso).

- (37) [In Grosz & Sidner' model] the notion of focus here reflects the (relative) salience of the entities in each focus space during the corresponding segment's portion of the discourse and thus refers to what is currently being talked about [...]. In fact, the notion of focus used in this AI-tradition is rather strongly reminiscent of the notion of topic [...]. (Zuo & Zuo 2001: 20)

Tale sovraestensione rispetto al modello originario non è priva di fondamento. Infatti, considerare il Topic del Discorso alla stregua di un'intenzione, come fanno Grosz & Sidner, solleva non pochi problemi, essendo l'intenzione un concetto non linguistico. Per contro, ammettere che vi sia, tra le altre, una componente intenzionale nel Topic del Discorso ben si coniuga con la sua natura interattiva e negoziabile: ammettere nel computo del Topic del Discorso non solo ciò di cui si parla ma anche il *perché* se ne parla potrebbe consentire una maggiore comprensione dei meccanismi di negoziazione. Inoltre, già a prima vista, la nozione di focus of attention, che riflette la entità salienti in ogni data porzione di Discorso, sembra corrispondere a ciò che è oggetto di Discorso, ossia al Topic del Discorso.

L'affermazione di Grosz & Sidner riportata in (36) comprende, oltre all'equiparazione del Topic del Discorso al DSP, anche l'equiparazione del Topic di frase alla nozione di "center", elaborata dalla stessa Grosz in lavori precedenti sulla risoluzione della referenza (Grosz *et al.* 1983). Per ragioni di completezza, nel prossimo paragrafo accennerò brevemente agli sviluppi che la nozione di center ha trovato nella letteratura.

3.1.2. La Centering Theory

La *Centering Theory* è un modello per la risoluzione della referenza che si fa tradizionalmente risalire a Grosz *et al.* 1983. Esso ha subito notevoli revisioni e ampliamenti (Grosz *et al.* 1995) ed è molto in voga nella linguistica computazionale contemporanea.³³ Secondo tale modello, certe entità menzionate nell'enunciato sono più "centrali" di altre e tale caratteristica di centralità impone condizioni sull'uso, da parte del parlante, delle espressioni referenziali.

Delineerò molto schematicamente i capisaldi della Centering Theory. Ad ogni enunciato U in un segmento di Discorso DS è associato un insieme di *forward looking centers* $C_f(U, DS)$; ad ogni enunciato diverso dall'enunciato che inizia un DS è assegnato un singolo *backward looking center* $C_b(U, DS)$. Il backward looking center di un enunciato U_{n+1} si connette ad uno dei forward looking centers di U_n . Nell'insieme dei forward looking di U_n esiste una scala di probabilità che determina i candidati favoriti a rappresentare il backward center di U_{n+1} . L'esempio seguente mostra una successione di quattro enunciati (segmenti di Discorso) e i rispettivi centri backward e forward looking.

³³ Si vedano ad esempio Walker *et al.* (1998), in cui compare anche il contributo di Barbara Di Eugenio sull'italiano, e Mitsakaki (2002), che passa in rassegna anche le numerose implementazioni del Centering.

- (38) 1. John has been having a lot of trouble arranging his vacation.
 2. He cannot find anyone to take over his responsibilities. (he = John)
 $C_b = \text{John}; C_f = \{\text{John}\}$
 3. He called up Mike yesterday to work out a plan. (he = John)
 $C_b = \text{John}; C_f = \{\text{John, Mike}\}$ (CONTINUE)
 4. Mike has annoyed him a lot recently.
 $C_b = \text{John}; C_f = \{\text{Mike, John}\}$ (RETAIN)
 e. He called John at 5 AM on Friday last week. (he = Mike)
 $C_b = \text{Mike}; C_f = \{\text{Mike, John}\}$ (SHIFT)

Ad un'analisi profonda, l'insieme dei centri forward looking rappresenta un insieme di Topic potenziali e la nozione di backward looking center è assimilabile al Topic di frase. Dunque, il modello intenzionale/attenzione di Grosz & Sidner (1986) e la Centering Theory non sono alternativi, quanto piuttosto complementari (come esplicitamente sostenuto in Grosz *et al.* 1995): nel primo, l'operazione di *Focusing* individua il Topic del Discorso; nel secondo, l'operazione di *Centering* individua il Topic di frase. In pratica, il Centering si occupa dell'*attentional state* di un segmento del Discorso e presiede alla coerenza locale; il Focusing si occupa dell'*attentional state* inter-segmentale e presiede alla coerenza globale. Non è un caso che Barbara Grosz abbia dato il contributo decisivo all'elaborazione di entrambi i modelli.³⁴

4. Implementazioni

Secondo Manning & Schützte (1999), lo studio del Topic del Discorso in linguistica computazionale (intesa dagli autori come linguistica computazionale "applicata", ovvero concretamente implementata) si articola in due aree: da un lato vi è il cosiddetto *Information Retrieval*, ossia la ricerca del Topic, dall'altro lato vi è la *Topic segmentation* o *Text Tiling*, ossia la segmentazione del testo in unità topicali. Entrambi i settori riscuotono notevole interesse al giorno d'oggi, date le molteplici applicazioni pratiche che possono trovare.³⁵ Le implementazioni non hanno alle spalle una sola teoria, né orientata teoricamente e neppure orientata computazionalmente. Piuttosto, esse assumono nozioni teoriche provenienti da più modelli, tra cui soprattutto i lavori di Discourse Analysis (*in primis* la RST, ma anche la RDA, più orientata computazionalmente fin dalle origini). Va notato però che, rispetto ai modelli originari da cui traggono spunti, i modelli implementati

³⁴ A *latere*, segnalo che sarebbe interessante il modello del Centering con la nozione di logoforicità.

³⁵ Le applicazioni dell'Information Retrieval e della Text Segmentation sono numerosissime: si pensi al possibile uso nelle biblioteche digitali, negli ipertesti, e, più in generale, in tutti i casi in cui si voglia estrarre informazioni da basi di dati. Ad esempio, nella seconda metà degli anni Novanta, l'agenzia americana di *Defence Advance Research Project* ha sviluppato un progetto denominato *Topic Detection and Tracking* per l'identificazione dei Topic e dei loro sviluppi. Lo scopo era quello di segmentare la stringa delle news di lingua americana e di lingua cinese (mandarino), sia scritte sia parlate, in storie individuali. Il progetto ha visto la partecipazione di più computazionalisti e la maggioranza degli algoritmi usati si basavano proprio sull'uso lessicale (per i riferimenti bibliografici si veda Tür *et al.* 2001).

operano drastiche semplificazioni, eliminando molti aspetti teorici ritenuti superflui oppure non trattabili da un sistema informatico.

Nelle implementazioni, sia di Information Retrieval sia di TextTiling, spesso il problema dell'identificazione del Topic del Discorso viene ridotto allo studio del *Topic shift*; in altre parole, invece di creare un algoritmo che identifichi i Topic del Discorso, si creano algoritmi per individuare il passaggio da un Topic all'altro, secondo un'intuizione che si trova già in Brown & Yule (1983, citazione al punto 4). L'idea base di questi sistemi consiste nell'identificare le parti del testo dove il vocabolario passa da un sub-Topic ad un altro, intendendo il Topic come una regione caratterizzata da *patterns* di co-occorrenza lessicale. A tal fine, molta attenzione è rivolta a quegli elementi che possono funzionare da indicatori (*cues*) del cambiamento di Topic: sono stati proposti modelli di identificazione del Topic basati su cues prosodici (Grosz & Hirshberg 1992, Hirshberg & Nakatani 1996), su cues discorsivi, tra cui i cosiddetti *Discourse Markers* (Passoneau & Litman 1997), su cues lessicali (Hearst 1997) o combinazioni di cues lessicali e prosodici (Tür *et al.* 2001). Nel paragrafo seguente, illustrerò un algoritmo per l'identificazione del Topic del Discorso basato su cues di tipo lessicale.

4.1. L'algoritmo di TextTiling di Marti Hearst

In Hearst (1997) viene proposto un sistema implementato di segmentazione del testo in paragrafi caratterizzati dal medesimo Topic.³⁶ L'algoritmo è strutturato in tre fasi: (1) *tokenization*; (2) *lexical score determination*; (3) *boundary identification*.

(1) Il processo di tokenizzazione consiste nella divisione del testo in input in unità lessicali individuali. Anziché suddividere la stringa di testo in frasi reali determinate con un *parsing* sintattico, l'algoritmo esegue una suddivisione lineare in pseudo-frasi di taglia predefinita w , definite *token-sequence*. La tokenizzazione serve a consentire il confronto tra unità di taglia uguale, in quanto il numero di termini condivisi nel caso di due frasi lunghe o nel caso di una frase corta e una lunga porterebbe punteggi incommensurabili. La tokenizzazione è sensibile al formato del testo in input. Ad esempio, se il testo veicola informazioni di *markup*, l'algoritmo salta le informazioni ausiliarie fino a localizzare il corpo del testo. Inoltre, l'algoritmo è provvisto di una *stop-list* composta da *stop-words*, ossia da parole funzionali, inutili per la ricerca del Topic, che non vengono considerate nella tokenizzazione. In un secondo passaggio, la token-sequenza viene ridotta alle sue radici morfologiche da una funzione di analisi morfologica basata su un lavoro di Karttunen, Koskenniemi & Kaplan (1987), che converte nomi flessi e verbi alle loro radici.³⁷

(2) Il processo di determinazione del punteggio lessicale può avvenire seguendo varie metodi matematici. Hearst (1997) ne esplora due. Il primo metodo, definito *block comparison*, confronta blocchi di testo adiacenti per vedere quanto siano simili in relazione al numero di parole in comune. Il secondo metodo, definito *vocabulary introduction*, confronta token-sequences per vedere quanto siano differenti in relazione al numero di parole nuove che compaiono. In questa sede mi concentrerò nell'illustrazione del metodo a comparazione di blocchi. Il primo passo consiste nel raggruppare le token-sequences in blocchi di taglia k . Tale valore k è stabilito tentando di approssimare la lunghezza media dei

³⁶ Si veda Pevzner & Hearst (2002) per un aggiornamento del sistema di metrica dell'algoritmo.

³⁷ Per un approfondimento sul processo di tokenizzazione, si veda Guo (1997).

paragrafi. Così come si erano utilizzate pseudo-frasi (token-sequences) invece delle frasi reali, ora si evita di utilizzare i reali paragrafi, poiché la loro lunghezza può essere molto irregolare e portare a confronti sbilanciati. Poi, ad ogni gap i tra due blocchi adiacenti viene assegnato un valore di similarità, che corrisponde alla sommatoria dei singoli *pesi* di ogni termine, ossia alla sommatoria delle singole frequenze con cui ciascun termine compare in ciascuno dei due blocchi. Per ottenere risultati confrontabili, si applica un processo di normalizzazione, in maniera da ricondurre i risultati a valori tra 0 e 1 compresi.

$$(39) \quad \text{lexical score}(i) = \frac{\sum_t w_{t,b1} w_{t,b2}}{\sqrt{\sum_t w_{t,b1}^2 \sum_t w_{t,b2}^2}}$$

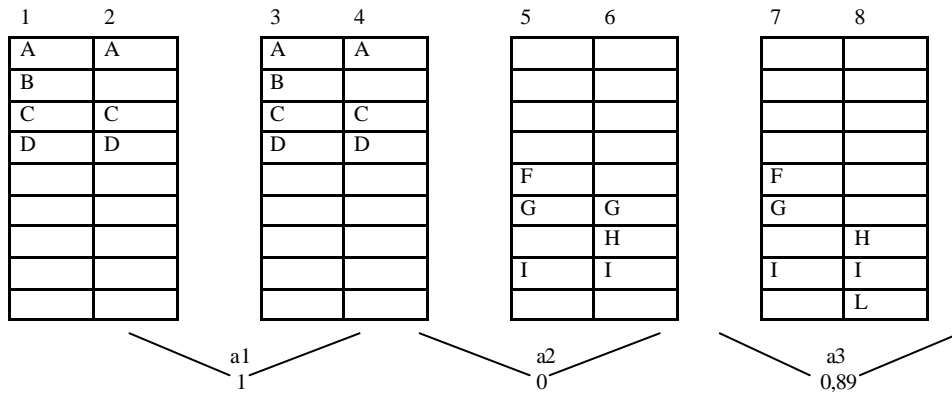
(3) La fase di boundary identification comprende un processo di assegnazione di depth score per ogni gap a partire dal lexical score del gap considerato in rapporto ai gap precedente e successivo. In altre parole, dati i punteggi lessicali y relativi a tre gap consecutivi a_1, a_2, a_3 , si calcola la “profondità della valle” che si registra a livello del gap intermedio a_2 . Al valore 0 corrisponde l’assenza di profondità, ossia l’assenza di discontinuità lessicale; al valore 2 la profondità è massima, ossia vi è totale discontinuità lessicale.

$$(40) \quad \text{depth score} = (y_{a1} - y_{a2}) + (y_{a3} - y_{a2})$$

Per illustrare l’algoritmo di Hearst, ho costruito un esempio *ad hoc*, poco probabile nella produzione spontanea, ma utile alla comprensione. Assegniamo alla taglia w delle pseudo-frasi il valore 4 (cioè ogni pseudo-frase è composta da 4 parole) e alla taglia k dei blocchi il valore 2 (cioè ogni blocco è composto da 2 pseudo-frasi). Ogni pseudo-frase è raffigurata da una colonna, dove le lettere maiuscole indicano gli *items* lessicali della frase; il raggruppamento di due colonne costituisce un blocco. L’esempio prevede che i primi due blocchi siano uguali, ovvero contengano gli stessi item lessicali lo stesso numero di volte; il punteggio lessicale del gap a_1 tra i primi due blocchi sarà 1. Il secondo ed il terzo blocco sono invece totalmente differenti, non avendo nessuna parola in comune: il punteggio lessicale del gap a_2 è 0. Il terzo ed il quarto blocco sono molto simili, ma differiscono riguardo a G (che compare due volte nel primo blocco e una sola volta nel secondo) e riguardo a L (che compare soltanto nel secondo blocco): il gap a_3 totalizza il punteggio di 0,89. A questo punto possiamo costruire un grafico che colloca in ascissa i gap e in ordinata i loro punteggi lessicali. Già guardando l’andamento del grafico, si nota una netta discontinuità a livello del gap a_2 . Applicando poi la formula per il calcolo del depth score, si ottiene per a_2 il valore di 1,89, che corrisponde effettivamente ad una discontinuità lessicale piuttosto alta.³⁸

³⁸ Si noti che la formula per il calcolo del depth score si applica a tre valori, ossia ai valori di punteggio lessicale relativi a tre gap consecutivi, e determina il depth score relativo al gap intermedio, in questo caso a_2 . Per la determinazione del depth score a livello del gap a_1 servirebbe il punteggio lessicale del blocco precedente, così come per a_3 servirebbe il punteggio lessicale del blocco successivo.

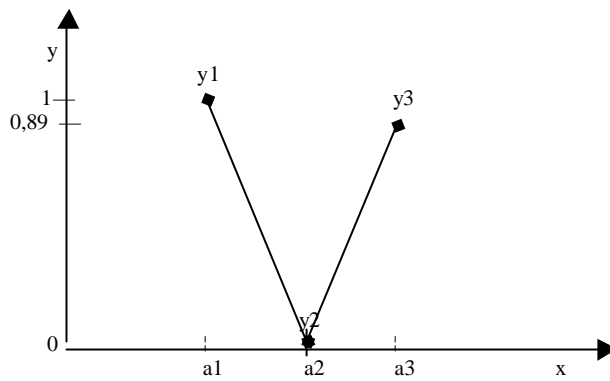
(41) $w = 4$
 $k = 2$



$$\text{score (a1)} = \frac{(2 \cdot 2) + (1 \cdot 1) + (2 \cdot 2) + (2 \cdot 2)}{\sqrt{(2^2 + 1^2 + 2^2 + 2^2) \cdot (2^2 + 1^2 + 2^2 + 2^2)}} = \frac{13}{\sqrt{169}} = \frac{13}{13} = 1$$

$$\text{score (a2)} = \frac{(2 \cdot 0) + (1 \cdot 0) + (2 \cdot 0) + (2 \cdot 0) + (0 \cdot 1) + (0 \cdot 2) + (0 \cdot 1) + (0 \cdot 1) + (0 \cdot 2)}{\sqrt{(2^2 + 1^2 + 2^2 + 2^2 + 0 + 0 + 0 + 0) \cdot (0 + 0 + 0 + 0 + 1^2 + 2^2 + 1^2 + 1^2 + 2^2)}} = 0$$

$$\text{score (a3)} = \frac{(1 \cdot 1) + (2 \cdot 1) + (1 \cdot 1) + (2 \cdot 2) + (0 \cdot 1)}{\sqrt{(1^2 + 2^2 + 1^2 + 2^2 + 0) \cdot (1^2 + 1^2 + 1^2 + 2^2 + 1^2)}} = \frac{8}{\sqrt{80}} = \frac{8}{9} = 0,89$$



$$\text{depth score} = (1 - 0) + (0,89 - 0) = 1 + 0,89 = 1,89$$

5. Confronti

I modelli considerati nelle sezioni precedenti muovono da punti di vista talmente differenti che un confronto che sia allo stesso tempo puntuale e complessivo risulta impraticabile. Infatti, mentre i tre modelli di Pragmatica Formale / Semantica del Discorso hanno aspetti semantici sottili, i modelli orientati computazionalmente eludono del tutto le questioni semantiche e si soffermano maggiormente sulla strutturazione del Discorso in segmenti. Tra i modelli computazionali, bisogna poi distinguere la proposta a stampo teorico di Grosz & Sidner (1986), dove la segmentazione del Discorso comprende nozioni psicologiche e intenzionali, dalla proposta implementata di Hearst (1997), dove la segmentazione è a base puramente lessicale. Un'operazione di confronto puntuale può essere praticata soltanto tra modelli simili nelle assunzioni fondamentali. Nel paragrafo seguente, procederemo dunque ad un confronto puntuale limitato ai modelli formali di stampo teorico esaminati nella sezione 2. Nel paragrafo 5.2, tenterò invece un confronto "sul campo" comprendente tutti i modelli considerati nella trattazione presente, che verranno applicati per sommi capi ad uno stesso frammento di testo. Ne emergerà un'estrema disparità di risultati, ossia verranno individuati Topic del Discorso diversi per lo stesso testo, nessuno dei quali pienamente soddisfacente. Pertanto, nel paragrafo 5.3, concluderò delineando alcuni punti che dovranno essere tenuti presenti nella costruzione di un modello efficace per il computo semantico-pragmatico del Topic del Discorso e, più in generale, per l'elaborazione di un modello della Struttura Informazionale del Discorso. Al fondo dell'articolo, verrà fornita una tabella riepilogativa che confronta i modelli considerati secondo i parametri ritenuti da me più rilevanti.

5.1. Confronto tra i modelli di Pragmatica Formale / Semantica del Discorso

Tutte le proposte considerate nella sezione 2 hanno alle spalle un modello complessivo del Discorso, che verte sulla Struttura Informazionale (come in van Kuppevelt e Roberts) oppure che contempla la Struttura Informazionale come una delle sue componenti (è il caso della SDRT). Tutti le proposte in considerazione assumono il ruolo strutturante del Topic, il quale struttura aggiornando, consentendo l'*updating*. Dunque, in tutte è adottata una prospettiva dinamica nell'analisi del Discorso e del Topic. Molto spesso, infatti, oltre alle espressioni "Pragmatica Formale" e "Semantica del Discorso", si è parlato di *Dynamic Semantics*. Tale denominazione indica ogni semantica che *treats the denotation of a sentence as its ability to update a discourse model* (McNally 1998: 164), dove per modello del Discorso si intende un modello della presunta conoscenza condivisa tra i partecipanti oppure, in altri approcci, il modello che il parlante ha del magazzino di conoscenze dell'ascoltatore.

Ciò che differenzia i vari lavori è il meccanismo di updating. Da un lato, vi sono i modelli di van Kuppevelt e di Roberts, con un approccio questioning dove l'updating avviene attraverso la modulazione di domande (esplicite o implicite) e risposte. Dall'altro lato, vi è la SDRT (e suoi derivati), dove l'updating sfrutta relazioni retoriche oppure relazioni topicali. Tuttavia recentemente anche la SDRT si sta adeguando all'approccio questioning (si vedano i lavori sulla SDRT dialogica in Asher & Lascarides 1998). Complessivamente, quindi, si ha un panorama in cui il *trend* è l'approccio questioning, secondo cui il meccanismo di strutturazione e al tempo stesso di aggiornamento del

Discorso si manifesta attraverso la modulazione di domande che esprimono Topic (del/nel Discorso). Nel prossimo paragrafo, formulerò alcune riflessioni specifiche sul questioning, che sembra riscuotere molto successo al giorno d'oggi, ma a proposito del quale numerose cautele devono essere assunte.

5.1.1. Valutazione del questioning

Prima di tutto, bisogna notare che ci sono differenze all'interno dell'approccio questioning, per quanto riguarda la denotazione del Topic-domanda. Secondo van Kuppevelt, il Topic è un termine della domanda, mentre secondo Roberts è la domanda intera. In termini semantici, il Topic per van Kuppevelt è l'intensione (o insieme di valori, o dominio di variabili) di un singolo termine, mentre per Roberts è l'intensione (o insieme di valori, o dominio di variabili) di una proposizione. Ad esempio, la domanda *Where did Robin go?*, in un modello di Discorso in cui vi siano soltanto Robin, lo zoo e l'acquario, rappresenta un Topic inteso, secondo van Kuppevelt, come l'insieme dei termini *{the zoo, the aquarium}*, e secondo Roberts, come l'insieme delle proposizioni *{Robin went to the zoo, Robin went to the aquarium}*. Anche se le proposizioni che costituiscono l'insieme-denotazione del Topic di Roberts si differenziano tra loro per un termine soltanto, facendo sì che l'approccio di Roberts e quello di van Kuppevelt, di fatto, si avvicinino, resta marcatissima la divergenza tra il considerare il Topic come un termine oppure una proposizione. D'altro canto, si è visto come la questione dell' "estensione" del Topic (un termine oppure una proposizione?) sia rilevante anche nella SDRT di Asher, dove la definizione riconosce il Topic come una proposizione ma l'esemplificazione porta all'individuazione di referenti topicali individuali.

Senza dubbio l'approccio questioning ha il vantaggio di fornire una semantica per il Topic, che, nella letteratura è stato indagato principalmente nei suoi aspetti sintattici e pragmatico/contestuale, tralasciando gli aspetti semantici. L'approccio questioning nasce all'interno della Pragmatica Formale per rispondere all'esigenza di una caratterizzazione semantica del Topic. Esso si fonda sull'equiparazione del Topic ad una domanda e sviluppa una teoria semantica della domanda che si rielabora la teoria classica di Hamblin (1973), ampliandola al contesto del Discorso.³⁹ In altre parole, l'approccio questioning offre una rappresentazione semantica del Topic di tipo "analitico", sciogliendo il Topic-domanda nelle sue risposte contestualmente accettabili.

Considerare il Topic come una domanda avente per denotazione un insieme di valori-risposte alternativi determinati contestualmente è senz'altro utile nello studio della Struttura Informazionale di alcuni tipi di Discorso, ad esempio i dialoghi task-oriented, dove sia il Topic globale dello scambio, inteso come domanda globale, sia i sotto-Topic (intesi come sotto-domande più specifiche) sono già stabiliti e l'insieme delle possibili risposte è per definizione ridotto. Non a caso, i modelli considerati in questo articolo sono stati esemplificati (e implementati) proprio su dialoghi task-oriented. Tuttavia, nel caso del dialogo spontaneo, le domande spesso non sono prevedibili, sono "di ampio respiro" e possono dar luogo ad un numero tendente all'infinito di risposte accettabili. Senza prendere l'esempio estremo di una domanda del tipo "Come va?" posta in apertura di conversazione, si consideri il caso di un dialogo sulle trascorse vacanze in cui, ad un certo punto, uno dei

³⁹ Una lucida rassegna sulle interpretazioni semantiche e pragmatiche della domanda si trova in Groenendijk & Stokhof (1997).

partecipanti chiede all'altro: "Per il resto, come sei stato?". In casi come questi, determinare la semantica del Topic-domanda significherebbe computare tutti i possibili valori contestualmente accettabili come risposta. A patto che una simile operazione sia possibile, si otterrebbe un insieme quasi-infinito, poco maneggevole e oltretutto piuttosto contro-intuitivo rispetto all'idea di "ciò dicui si parla".

Un'ulteriore obiezione all'approccio questioning deriva dal fatto che la domanda-Topic non sempre può contenere già, all'interno della sua intensione semantica, la sua risposta. Bisogna infatti considerare i processi di negoziazione del Topic, per cui esso va costruendosi e rinnovandosi ad ogni stadio della conversazione, grazie ai vari contributi dei partecipanti. In altre parole, un partecipante può dare una risposta contestualmente imprevedibile, che non era compresa nell'intensione del Topic-domanda, eppure efficace.

Tra l'altro, si noti che l'idea di un insieme di valori alternativi che rappresentano la denotazione, l'interpretazione semantica, di un termine è stata usata anche nell'ambito degli studi sul Focus. Vi è un filone di ricerca, che si fa risalire a Rooth (1992) e si definisce di *Alternative Semantics*, che interpreta il Focus in maniera molto simile a come è trattato il Topic dall'approccio questioning, ossia come un insieme dei possibili valori. Anche l'approccio questioning al Topic potrebbe dunque a pieno titolo definirsi di tipo *Alternative Semantics*. Mettendo insieme l'approccio di *Alternative Semantics* al Focus e l'approccio questioning al Topic, ne deriva una perdita di specificità sia del Focus sia del Topic, che vengono ad avere la medesima interpretazione semantica. Come fa notare Kadmon, la denotazione della risposta-Focus deve coincidere con la denotazione della domanda-Topic.

- (42) [...] according to Hamblin (1973), a question denotes the set of propositions that constitute all the (partial) answer to that question –true and false answer alike. Given this, a simple constraint forcing question-answer congruence can be stated as follow: the focus semantic value of the answer (a set of alternative proposition) must be identical to the denotation of the question (also a set of alternative proposition). (Kadmon 2001: 292)

In tale quadro, la differenza tra Topic e Focus starebbe soltanto nel loro valore pragmatico, mentre, dal punto di vista vero-condizionale coinciderebbero. Si ritornerebbe così a modelli, ormai superati, in cui la distinzione tra semantica e pragmatica è netta, e la pragmatica è responsabile di fenomeni che vanno al di là della vero-condizionalità. Il rapporto tra approccio questioning al Topic e approccio di *Alternative Semantics* al Focus andrebbe meglio indagato nelle sue implicazioni semantiche.⁴⁰

Il punto più debole dell'approccio questioning al Topic resta quello delle domande implicite. van Kuppevelt sostiene che sia sempre possibile, tramite un algoritmo non meglio definito, ricavare per ogni asserzione una domanda implicita a cui dare il valore di Topic. Roberts afferma che, perlomeno per la lingua inglese, il Focus prosodico può aiutare a determinare la domanda implicita. Manca però un algoritmo, funzionante e convincente, per ricavare le presunte domande implicite di cui ogni Discorso, anche quello monologico, dovrebbe strutturarsi. Il problema non sta tanto nell'ammettere che il Topic del Discorso si componga di materiale non enunciato esplicitamente (cosa che, come accennato in 1.1 è

⁴⁰ Peraltro, spesso nei modelli proposti in Pragmatica Formale, l'approccio questioning al Topic e l'approccio di *Alternative Semantics* al Focus coesistono, rappresentando due "moduli" interagenti della Struttura Informazionale complessiva. Così è nel caso del modello di Roberts, ma si vedano anche von Stechow (1994) e Büring (1997). Alcune osservazioni sull'esito dei due approcci si trovano in Portner & Yabushita (1998) e nel capitolo dedicato al Focus in Kadmon (2001).

altamente probabile), quanto piuttosto nell'ammere che il Discorso si strutturi in domande, anche laddove esse non sono formulate. Personalmente, sembra più naturale ammettere che lo sviluppo del Discorso non sia necessariamente legato ad un processo di questioning, soprattutto in considerazione delle tipologie discorsive non quesry-based e non task-oriented.

Come ultima osservazione, segnalo che l'approccio questioning modella il Discorso secondo una struttura domanda-risposta che procede da domande-Topic di portata superiore a sotto-domande-Topic. Infatti, il Discorso viene modellato sfruttando meccanismi a pila, ossia di tipo *push-down*. Nuovamente, una rappresentazione di questo tipo funziona bene per dialoghi task-oriented che procedano dallo scopo globale allo scopo particolare. Tuttavia sembra improbabile che un algoritmo push-down possa rendere conto in maniera efficace dei molti casi di andamento cataforico del Topic del Discorso.

5.2. Inapplicabilità pratica

In questo paragrafo, vorrei evidenziare come nessuno dei modelli considerati sia in grado di monitorare effettivamente la costruzione del Topic del Discorso in un caso di dialogo reale quale quello rappresentato in (1). Una prima considerazione da fare riguarda la difficoltà di testare i vari modelli su uno stesso testo, dal momento che ciascuna delle proposte considerate qui è stata esemplificata dagli autori su ben definite tipologie testuali e non su altre. Ad esempio, i modelli che adottano il questioning sono illustrati soltanto su dialoghi task-oriented, mentre la SDRT è illustrata soltanto su monologhi. Dunque, l'operazione di confronto applicato che compirò in questo paragrafo è per certi versi arbitraria, nel senso che mancano precisi riferimenti nei modelli originari presi in considerazione. Tale operazione è tuttavia necessaria nel momento in cui si voglia testare l'effettivo funzionamento delle varie proposte interpretative.

Secondo van Kuppevelt, il Topic del Discorso dovrebbe essere "hai fatto un corso di acquerelli?", poiché da esso hanno origine tutte le sottodomande e le asserzioni successive (e si potrebbe ipotizzare come feeder l'arrivo, sulla scena della conversazione, di KAT, che i partecipanti fanno essere di ritorno da un corso di acquerelli). Nella proposta di Roberts il Discourse Topic non è esplicitamente formulato; tuttavia, data l'organizzazione gerarchica delle QUD, si può pensare che il Topic del Discorso corrisponda alla domanda più "bassa" nella struttura a pila. Dunque, nel caso dell'esempio (1), il Topic del Discorso corrisponderebbe nuovamente a "hai fatto un corso di acquerelli?", che avrebbe come interpretazione semantica l'insieme composto dalle proposizioni {K ha fatto un corso di acquerelli, K non ha fatto un corso di acquerelli}. Eppure, intuitivamente, il Topic del Discorso del testo in (1), sembra essere più articolato della domanda "hai fatto un corso di acquerelli" e soprattutto il suo valore semantico, il suo significato, sembra essere molto diverso dalle due alternative {K ha fatto un corso di acquerelli, K non ha fatto un corso di acquerelli}. Esso sembra piuttosto comporsi anche dell'idea di "guardare gli acquerelli" e dell'idea di "non sporcare gli acquerelli"; in altre, parole, non sembra limitato ad una domanda specifica, ma appare come un complesso costruito attingendo a tutti gli enunciati del Discorso.

Passando alla proposta di Asher, assumendo come guida l'esempio considerato in (34) e non la definizione teorica (secondo cui il Topic è una proposizione), bisognerebbe estrarre dal testo il referente "acquerelli". L'idea che il referente del Discorso "acquerelli" rappresenti il Topic del Discorso in (1) è sicuramente più corrispondente all'intuizione.

Tuttavia, non è chiaro su quale base linguistica possa avvenire l'estrazione del referente "acquerelli", e non piuttosto di altri referenti del Discorso, i quali giocano comunque un ruolo nella determinazione del Topic.

Per quanto riguarda la proposta di Grosz e Sidner, seguendo la versione "vulgata" del modello e identificando i Topic del Discorso con i Focus space, dovremmo postulare, per il testo 1, una serie di segmenti, ciascuno dei quali dotato di un proprio Focus space. Probabilmente, il primo segmento risulterebbe composto dagli enunciati 1 2, e corrisponderebbe ad un Focus space composto dei referenti "corso di acquerelli", "ultima lezione", oltre che di due intenzioni di tipo linguistico (una per ciascun parlante) che mi è difficile formulare. Non mi cimenterò nell'operazione di segmentazione secondo il modello attenzionale/intenzionale di Grosz & Sidner, limitandomi a precisare la sua poca maneggiabilità pratica.

Infine, secondo l'algoritmo di Hearst, dovremmo calcolare in termini matematici, la co-occorrenza lessicale del testo in (1). Probabilmente, se considerassimo il Discorso complessivo da cui (1) è stato tratto, noteremmo una discontinuità lessicale tra il segmento di testo precedente ad (1) ed (1), a testimonianza del fatto che (1) inaugura un nuovo Topic del Discorso. Tuttavia, l'algoritmo di Hearst individua il punto di Topic shift, ma non determina in alcun modo il Topic in sé. Bisogna però considerare che simili algoritmi a base lessicale, per essere davvero funzionanti e monitorare efficacemente lo snodarsi dei Topic, dovrebbero essere ulteriormente affinati. In particolare, devono poter riconoscere possibili digressioni momentanee, come l'enunciato 5 nell'esempio (1), che, se computate indistintamente, porterebbero ad individuare falsi Topic-shift. Inoltre, tali algoritmi dovrebbero disporre di reti semantiche in grado di individuare quando effettivamente un cambiamento nella distribuzione lessicale corrisponde ad un Topic shift. In altre parole, se scompare la parola "acquerelli", non necessariamente è cambiato il Topic del Discorso, che può essersi specificato in "pittura" o "colori", ecc.

5.3. Alcune acquisizioni

Ciò che mi propongo di elaborare in futuro è un algoritmo efficace per il computo del significato del Topic del Discorso (significato pragmatico e non vero-condizionale, s'intende). Tale algoritmo dovrebbe andare ad integrare il Modello di Struttura Informazionale Estesa delineato in 1.1. Finora, la natura "estesa" del Modello si riferisce all'ampliamento del bagaglio teorico (che, come si è visto, ingloba nozioni non strettamente informative, provenienti dalla teoria degli atti linguistici e dalla pragmatica griceana); in futuro, vorrei che il Modello venisse ampliato anche nel senso dell'unità di misura, passando dalla dimensione dell'enunciato alla dimensione del Discorso, per rendere conto del il flusso di informazione nel Discorso parlato. Le intuizioni di base da cui muoverò nell'elaborazione del Modello sono sostanzialmente due e sono state in parte già delineate in 1.1: la prima è che il ruolo guida nella strutturazione informazionale del Discorso spetta al Topic del Discorso; la seconda è che la Relevance intra-discorsiva (che lega i Comment di enunciato ai Topic del Discorso) svolga un ruolo strutturante nella costruzione del Discorso stesso. Nella letteratura esistente, sebbene manchi un modello applicabile efficacemente per il computo del Topic del Discorso, ho riscontrato numerose conferme alle mie intuizioni di partenza.

In particolare, ho trovato conferme sull'importanza di attribuire al Topic il ruolo essenziale nella costituzione della Struttura Informazionale a livello di Discorso. Già nella

mia tesi di laurea, ero giunta alla conclusione che la Struttura Informazionale di enunciato verte sul Comment, mentre, salendo a livello di Discorso, il ruolo guida passa al Topic, inteso come Topic del Discorso. In altre parole, a livello di enunciato non si può fare a meno del Comment (ma si può fare a meno del topic), mentre, a livello di Discorso non si può fare a meno del Topic. Infatti, tutti i modelli considerati, che sono modelli del Discorso e non della frase, si concentrano non sul ruolo del Comment bensì sul ruolo del Topic (inteso talvolta come Topic del Discorso talvolta come Topic di frase nel Discorso).

Inoltre, ho visto confermata l'importanza della nozione di "Relevance" nella gestione del Topic Management. Nella mia tesi avevo interpretato in termini di Relevance intra-enunciato la relazione di predicazione semantico-pragmatica che si dà tra Topic e Comment. Ampliando la prospettiva al Discorso, e considerando la Relevance ricorsivamente, ho postulato un legame di Pertinenza anche tra Comment e Topic del Discorso. Nella rassegna condotta nel presente articolo, la Relevance è comparsa varie volte, in vesti differenti ma conservando la sua natura di relazione di pertinenza semantico-pragmatica e configurandosi come il più probabile candidato al ruolo di "collante" tra le unità informazionali del Discorso (si veda la voce "nozione di Relevance" nella tabella al fondo dell'articolo). Si è visto come, nei modelli di questioning, la Relevance colleghi tra loro non soltanto domande e risposte ma anche domande e sotto-domande. Da questa osservazione traggio un'idea da sviluppare in futuro: che i Topic di enunciato, qualora presenti, si leghino ai Topic di Discorso attraverso una ulteriore variante della Relevance griceana.

Infine, più in generale, ho visto confermata la necessità di inquadrare lo studio della Struttura Informazionale in un modello complessivo del Discorso, avendo la Struttura Informazionale una funzione strutturante nel Discorso. Tuttavia, non credo che si debba giungere all'identificazione della struttura del Discorso con la Struttura Informazionale. Un modello del Discorso dovrebbe contemplare, mantenendole distinte ma analizzandone le interazioni, sia la Struttura Informazionale, sia altri aspetti strutturanti quali, ad esempio, la risoluzione delle referenze anaforiche e le configurazioni illocutive e perlocutive.⁴¹

Una volta definite le nozioni in gioco, il Modello di Struttura Informazionale Estesa che mi propongo di elaborare dovrà concentrarsi sulla costruzione di un algoritmo per il computo del significato pragmatico del Topic del Discorso, che, altrimenti, resterebbe una nozione fondata soltanto intuitivamente. Dall'esame della letteratura, è emerso che un elemento fondamentale nel computo del Topic del Discorso è rappresentato dal contesto. Infatti, il Topic del Discorso non si compone soltanto di esplicito materiale linguistico prelevato dagli enunciati del Discorso, ma comprende anche materiale non espresso esplicitamente e determinato contestualmente. Il problema della formalizzazione del Topic del Discorso si sposta dunque verso il problema della formalizzazione del contesto, una questione che è presente all'attenzione dei linguisti da almeno un ventennio. A ben vedere, il filone della Pragmatica Formale / Semantica del Discorso nasce precisamente dalla necessità di trattare formalmente il contesto/Discorso. Sono stati proposti svariati modelli per il trattamento formale del contesto, tra cui la *Discourse Representation Theory*, la *File Change Semantics*, ecc. Tuttavia, tali modelli si sono concentrati sui referenti del Discorso, intesi come sintagmi nominali. I vari modelli di Pragmatica Formale che abbiamo

⁴¹ È stata la prof. Lavinia Merlini Barbaresi a farmi notare la necessità di delimitare con cautela il ruolo della Struttura Informazionale, che, se esteso eccessivamente, finirebbe con l'assorbire la struttura del Discorso nel suo complesso, offuscando altri aspetti strutturanti, quali, ad esempio, l'andamento illocutivo e perlocutivo del testo.

esaminato nel corso di questo lavoro assumono sempre un dominio contestuale del Discorso D, dal quale si attinge per la determinazione del Topic. Tuttavia tale dominio contestuale D si compone di referenti individuali, in genere sintagmi nominali, e funziona per rendere conto del Topic di frase *nel* Discorso, ma non per spiegare il Topic *del* Discorso. Resta ancora del tutto da formalizzare il contesto nei suoi aspetti linguistici ma impliciti e nei suoi aspetti cognitivi di *shared knowledge* e di *belief* dei partecipanti, aspetti dai quali la costruzione semantico-pragmatica del Topic del Discorso crucialmente dipende.⁴²

Indirizzo dell'Autore:

Scuola Normale Superiore, Laboratorio di Linguistica
Piazza dei Cavalieri 7, 56100 Pisa, Italy
v.bambini@sns.it

Riferimenti bibliografici

- Akman, Varol & Carla Bazzanella, a cura di (2003), *Journal of Pragmatics 35, Special issue: Context*, Amsterdam, Elsevier.
- Asher, Nicholas (1993), *Reference to Abstract Objects in Discourse*, Dordrecht, Kluwer.
- Asher, Nicholas & Isabel Gómez Txurruka (1995), "Extending SDRT to integrate a theory of the informational partition", Technical report, *Institute for Logic Cognition, Language and Information*.
- Asher, Nicholas & A. Lascarides (1998), "Questions in dialogue", *Linguistics and Philosophy* 21: 237-309.
- Asher, Nicholas (1999), "Discourse structure and the logic of conversation", in Ken Turner (a cura di), 19-48.
- Bambini, Valentina (2001), *La Struttura Informativa dell'Enunciato*, Tesi di Laurea, Università di Pisa.
- Bartsch, R. (1976), "Topik-Focus-Struktur und kategoriale Syntax", in V. Ehrlich & P. Finke (a cura di), *Grammatik und Pragmatik*, Kronberg, Scriptor Verlag.
- de Beaugrande, Robert. A. & Wolfgang U. Dressler (1981), *Einführung in die Textlinguistik*, Tübingen, Max Niemeyer Verlag.
- van Benthem, Johan & Alice ter Meulen, a cura di (1997), *Handbook of Logic and Language*, Amsterdam, Elsevier.
- Berthoud, A. C. (1996), *Paroles à propos: approche énonciative et interactive du topic*, Paris, Ophrys.
- Bos, Johan (1995), "Anaphora resolution and discourse segmentation in Artificial Intelligence", in Frank van Eynde & Valerio Allegranza (a cura di), 175-189.

⁴² Sulla formalizzazione del contesto, segnalo l'articolo di Stalnaker (1998), che, sebbene di stampo prettamente filosofico, è ricco di spunti. Si veda poi il recentissimo numero speciale di *Journal of Pragmatics*, curato da Akman & Bazzanella (2003), dedicato al tema del *Context*, esaminato secondo vari approcci.

- Brown, Gillian & George Yule (1983), *Discourse Analysis*, Cambridge, Mass., Cambridge University Press.
- Büring, Daniel (1997), *The Meaning of Topic and Focus. The 59th Street Bridge Accent*, London, Routledge.
- Cresti, Emanuela (2000), *Corpus di Italiano Parlato*, vol. I: *Introduzione*, vol. II: *Campioni*, Firenze: Accademia della Crusca.
- Davis, Paul C. (2000), "Presupposition resolution with discourse information structures", in Jennifer S. Miller, Tsan Hung & Craige Roberts (a cura di), *Varia, Ohio State University Working Papers in Linguistics* 54: 25-58.
- Di Eugenio, Barbara (1998), "Centering in Italian", in Marilyn Walker *et al.* (a cura di), 115-137.
- Downing, Angela (1997), "Encapsulating discourse topic", *Estudios ingleses de la Universidad Complutense*, 5: 147-168.
- van Eijck, Jan & Hans Kamp (1997), "Representing discourse in context", in Johan van Benthem & Alice ter Meulen (a cura di), 179-237.
- Erteschik-Shir, Nomi (1997), *The Dynamics of Focus Structure*, Cambridge: Cambridge University Press.
- van Eynde, Frank & Valerio Allegranza (1995), *Semantics and Discourse: A NLP Perspective*, Brussel/Luxembourg: European Commission.
- von Fintel, Kai (1994), *Restriction on Quantifier Domain*, Ph.D. Dissertation, University of Massachusetts, Amherst.
- Giora, Rachel, Meiran, Nachshon, & Oref, Paz (1996), "Identification of written discourse topics by structure coherence and analogy strategy: general aspects and individual differences"; *Journal of Pragmatics*, 26(4): 455-474.
- Grabski, Michael (2000), "Satztopic und Diskurstopic in Elaboration-Kontexten", in K. Schwabe *et al.* (a cura di), *Issues on topic. Zentrum für allgemeine Sprachwissenschaft Papers in Linguistics*, 20: 173-207.
- Groenendijk, Jeroen & Martin Stokhof (1997), "Questions", in Johan van Benthem & Alice ter Meulen (a cura di), 1055-1124.
- Grosz, Barbara J. & Candace Sidner (1986), "Attention, intention and the structure of discourse", *Computational Linguistics* 12: 175-204.
- Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein (1983), "Providing a unified account of definite noun phrases in discourse", *Proceedings of the 21st Annual Meeting of the Association of Computational Linguistics*, 44-50.
- Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein (1995), "Centering: A framework for modelling local coherence of discourse", *Computational Linguistics* 21 (2): 203-226.
- Guo, Jin (1997), "Critical tokenization and its properties", *Computational Linguistics* 23(4).
- Hamblin, C. (1973), "Questions in Montague English", *Foundations of Language*, 10: 41-53.
- Hearst, Marti A. (1997), "TextTiling: Segmenting text into multi-paragraph subtopic passages", *Computational Linguistics* 23(1): 33-64.
- Hendriks, Herman L. W. (2002), "Links without locations. Information packaging: from cards to boxes", *Linguistische Berichte, Sonderheft* 10: 172-197.
- Hirschberg, Julia & Christine Nakatani (1998), "Acoustic indicators of Topic segmentation", in Robert H. Mannell & Jordi Robert-Ribes (a cura di), *Proceedings of*

- the International Conference on Spoken Language Processing*, 976-979, Sydney, December, Australian Speech Science and Technology Association.
- Kadmon, Nirit (2001), *Formal Pragmatics. Semantics, Pragmatics, Presupposition, and Focus*, Oxford, Blackwell.
- Kamp, Hans (1981), "A theory of truth and semantic representation", in Jeroen Groenendijk, Theo M. V. Janssen & Martin B. J. Stokhof (a cura di), *Formal Methods in the Study of Language, Proceeding of the Third Amsterdam Colloquium*, Amsterdam: Mathematisch Centrum, 277-321; ristampato in Jeroen Groenendijk, Theo M. V. Janssen & Martin B. J. Stokhof (a cura di), *Truth, Interpretation and Information. Selected Papers from the Third Amsterdam Colloquium*, Dordrecht, Foris, 1984, 1-41.
- Kamp, Hans & Uwe Reyle (1993), *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht, Kluwer Academic Publisher.
- Karttunen, Laurie (1976), "Discourse referents", in James MacCawley (a cura di), *Syntax and Semantics 7*, New York: Academic Press, 363-385.
- Karttunen, Lauri, Kimmo Koskeniemi & Ronald Kaplan (1987), "A compiler for two-level phonological rules", in Mary Dalrymple (a cura di), *Tools for Morphological Analysis*, Center for the Study of Language and Information, Stanford, CA.
- Keenan, Elinor O. & Bamby Schieffelin (1976), "Topic as a discourse notion", in Charles N. Li (a cura di), *Subject and Topic*, New York, Academic Publisher.
- Kintsch, W & T van Dijk (1978), "Towards a model of text comprehension and production", *Psychological Review* 85: 363-394.
- Klein, W. & C. Von Strutterheim (1987), "Quaestio und referentielle Bewegung in Erzählungen", *Linguistische Berichte* 109: 163-183.
- Kruijff-Korabayova, Ivana & Geert-Jan Kruijff (1997), "Topic-focus articulation in DRT", *Proceedings of the 11th Amsterdam Colloquium*, December, 17-20.
- Kruijff, Geert-Jan & Jan Schaake (1997), "Discerning relevant information in discourse using TFA", in Ruslan Mitkov & Nicolas Nicolov (a cura di), 259-270.
- van Kuppevelt, Jan (1991), *Topic en Comment: Expliciete en Impliciete Vraastelling in Discourse*, Ph.D. Dissertation, University of Nijmegen.
- van Kuppevelt, Jan (1994), "Directionality in Discourse", in Peter Bosh & Rob A. van der Sandt (a cura di), *Focus and Natural Language Processing, IBM Working Papers of the IBM Institute for Logic and Linguistics 7*, 3 voll., Heidelberg: IBM Deutschland, 485-501.
- van Kuppevelt, Jan (1995a), "Discourse structure, topicality and questioning", *Journal of Linguistics* 31: 109-147.
- van Kuppevelt, Jan (1995b), "Main Structure and side structure in discourse", *Linguistics* 33: 809-833.
- van Kuppevelt, Jan (1996), "Inferring from topics: scalar implicatures as topic-dependent", *Linguistics and Philosophy* 19 (4): 393-443.
- Mann, W. C. & Sandra A. Thompson (1988), "Rhetorical structure theory: Towards a functional theory of text organization", *Text* 8(3): 243-281.
- Manning, Christopher D. & Hinrich Schütze (1999), *Foundations of Statistical Natural Language Processing*, Cambridge, Mass.: MIT Press.
- Miltsakaki, Eleni (2002), "Towards an aposynthesis of Topic continuity and intrasentential anaphora", *Computational Linguistics* 28(3): 319-355.

- Mc Nally, Louise (1997), "On recent formal analyses of topic", in Jonathan Ginzburg (a cura di), *The Tbilisi Conference on Logic, Language, and Computation: Selected Papers*, Stanford, Center for Study of Language and Information, 149-162.
- Mc Nally, Louise (1998), "On the linguistic encoding of information packaging instructions", in Peter Culicover & Louise Mc Nally (a cura di), *Syntax and Semantics 29, The Limits of Syntax*, New York: Academic Press, 161-183.
- Mitkov, Ruslan & Nicolas Nicolov, a cura di (1997), *Recent advances in Natural Language Processing*, Amsterdam/Philadelphia, John Benjamins.
- Moore, Johanna D. & Martha E. Pollack (1992), "A problem for RST: The need for multi-level discourse analysis", *Computational Linguistics* 18(4): 537-544.
- Moser, Margaret G. & Johanna D. Moore (1996), "Towards a synthesis of two accounts of discourse structure", *Computational Linguistics* 22(3): 409-419.
- Muskens, Reinhard, Johan van Benthem & Albert Visser (1997), "Dynamics", in Johan van Benthem & Alice ter Meulen (a cura di), 587-648.
- Nomoto, Tadashi (1997), "Effects on grammatical annotation on a Topic identification task", in Ruslan Mitkov & Nicolas Nicolov (a cura di), 235-246.
- Van Oosten, Jeanne (1985), *The nature of subject, topic, and agent: a cognitive explanation*, Bloomington, Indiana University Linguistics Club.
- Passoneau, Rebecca J. & Diane J. Litman (1997), "Discourse segmentation by human and automated means", *Computational Linguistics* 23(1): 103-139.
- Peregrin, Jaroslav (1999), "The pragmatization of semantics", in Ken Turner (a cura di), 419-442.
- Pevzner, Lev & Marti A. Hearst (2002), "A critique and improvement of an evaluation metric for text segmentation", *Computational Linguistics* 28(1).
- Poesio, Massimo & Barbara Di Eugenio (2001), "Discourse structure and anaphoric accessibility", Paper presented at the ESSLI 2001 Workshop on *Information Structure, Discourse Structure and Discourse Semantics*.
- Prévoit, Laurent (2001), "Topic Structure in Route Explanation Dialogues", Paper presented at the ESSLI 2001 Workshop on *Information Structure, Discourse Structure and Discourse Semantics*.
- Ramm, Wiebke (1997), "Discourse constraints on theme selection", in Ruslan Mitkov & Nicolas Nicolov (a cura di), 247-258.
- Roberts, Craige (1996), "Information structure in discourse: towards an integrated formal theory of pragmatics", in Jae-Hak Yoon & Andreas Kathol (a cura di), *Papers in Semantics, Ohio State University Working Papers in Linguistics* 49, Columbus, Ohio State University, 91-136.
- Roberts, Craige (1998a), "Focus, the flow of information, and Universal Grammar", in Peter Culicover & Louise Mc Nally (a cura di), *Syntax and Semantics 29, The Limits of Syntax*, New York, Academic Press, 109-160.
- Roberts, Craige (1998b), "The place of centering in a general theory of anaphora resolution", in Marilyn A. Walker *et al.* (a cura di), 359-399.
- Rooth, M. (1992), "A theory of focus interpretation", *Natural Language Semantics* 1: 75-116.
- Stalnaker, Robert (1998), "On the representation of context", *Journal of Logic, Language and Information* 7: 3-19.
- Tracy, K. (1984), "Staying on topic: An explication of conversational relevance", *Discourse Processes* 7: 447-464.

- Tür, Gökhan, Dilek Hakkani-Tür, Andreas Stolke & Elizabeth Shriberg (2001), "Integrating prosodic and lexical cues for automatic Topic segmentation", *Computational Linguistics* 27(1): 31-57.
- Turner, Ken, a cura di (1999), *The Semantics/Pragmatics Interface from Different Points of View*, Amsterdam, Elsevier.
- Varile, Giovanni Battista & Antonio Zampolli, a cura di (1997), "Survey of the State of the art in Human Language Technology", *Linguistica Computazionale XII-XIII*, Pisa, Giardini Editori.
- Vennemann, Theo (1975), "Topics, sentence accent, ellipsis: a proposal for their formal treatment", in E. L. Keenan (a cura di), *Formal Semantics of Natural Language*, Cambridge, Cambridge University Press, 313-328.
- Walker, Marilyn A., Joshi, Aravind K., & Prince, Ellen F., a cura di (1998), *Centering Theory in Discourse*, Oxford, Clarendon Press.
- Walker, Marilyn A. (1996), "Limited attention and discourse structure", *Computational Linguistics* 22(2): 255-264.
- Walker, Marilyn A. & Johanna D. Moore (1997), "Empirical studies in Discourse", *Computational Linguistics* 23(1): 1-12.
- Xturruka, Isabel Gómez (1997), *The informational partition and discourse structure*, Ph.D. Thesis, University of the Basque Country, Donostia, Spagna.
- Zuo, Yan & Wei Zuo (2001), *The Computing of Discourse Focus*, München, Lincom Europa.

Tabella comparativa: I modelli di Pragmatica Formale

AUTORE	ORIENTAMENTO	FRAMEWORK	NOZIONE DI TOPIC DI FRASE	INTERPRETAZIONE SEMANTICA DEL TOPIC	NOZIONE DI DISCOURSE TOPIC	FUNZIONE DEL TOPIC NEL DISCORSO	NOZIONE DI RELEVANCE	TIPOLOGIA TESTUALE CONSIDERATA	IMPLEMENTAZIONI	TEST SU CORPORA
Van Kuppevelt 1991 e segg.	Pragmatica Formale; orientamento teorico	Modello dinamico del Discorso, articolato secondo la modulazione gerarchica della struttura Topic/Comment, realizzata attraverso un procedimento domanda/risposta (dove le domande possono essere sia esplicite sia implicite)	Ciò che è domandato (un termine della domanda), dunque l'insieme di una o più entità del Discorso che sono oggetto di domanda	Intensione del termine Topic, ossia l'insieme di possibili valori che il termine Topic può assumere nella situazione attuale (il Comment esprime il valore specifico assunto dal Topic nella situazione attuale, dunque la sua estensione)	Insieme dei Topic originati da uno stesso feeder; che può coincidere con un singolo Topic semplice	Strutturante, attraverso il processo gerarchico di domande/risposte	Sotto forma di <i>Principle of Recency</i> e <i>Dynamic Principle of Topic Termination</i> , che regolano la creazione e la terminazione delle domande-Topic	In linea teorica senza restrizioni, ma di fatto pochi esempi, tutti costruiti, di dialoghi <i>query-based</i>	∅	∅
Roberts 1996 e segg.	Pragmatica Formale; orientamento teorico	Teoria pragmatica del Discorso in cui il linguaggio è un gioco di domande intenzionali	Definito come <i>Question Under Discussion</i> (QUD), ossia la domanda in corso, da intendersi come una mossa <i>set-up</i> che definisce una meta del Discorso	Intensione della QUD; dunque, l'insieme di proposizioni che possono funzionare da risposte alla QUD	Non menzionato esplicitamente; potrebbe corrispondere e alla QUD più bassa della pila, l'ultima ad essere rimossa dall'automa	Strutturante, attraverso il processo di domande-risposte (dove le domande rappresentano mete del Discorso)	Regola che prescrive che ogni contributo costituisca o un'asserzione in risposta ad una esplicita o implicita domanda o una sotto-domanda della precedente QUD	Come sopra	Proposta in linea teorica da Roberts con il meccanismo della pila QUD-stack; realizzata da Davis (2000) per dialoghi uomo-macchina <i>task-oriented</i> (con Topic del Discorso stabilito)	∅

Tabella comparativa: I modelli di Semantica del Discorso

AUTORE	ORIENTAMENTO	FRAMEWORK	NOZIONE DI TOPIC (DI FRASE O ENUNCIATO)	INTERPRETAZIONE SEMANTICA DEL TOPIC	NOZIONE DI DISCOURSE TOPIC	FUNZIONE DEL TOPIC NEL DISCORSO	NOZIONE DI RELEVANCE	TIPOLOGIA TESTUALE CONSIDERATA	IMPLEMENTAZIONI	TEST SU CORPORA
Asher 1993	Semantica del Discorso ; orientamento teorico	Modello dinamico di rappresentazione semantica del Discorso definito <i>Segmented Discourse Representation Theory</i> , ossia <i>Discourse Representation Theory</i> (Kamp & Reule 1981) integrata con relazioni retoriche (tratte per lo più dalla RST di Mann & Thompson 1986)	Un costituente riassunto per un gruppo di costituenti (spesso si tratta di una proposizione)	Una <i>Discourse Representation Structure</i> , composta da referenti linguistici più condizioni (predicati) su di essi (raffigurata graficamente con un box)	∅	Strutturante, attraverso un processo di aggiornamento che può sfruttare due relazioni del discorso: <i>Elaboration</i> e <i>Continuation</i> (il <i>topic-based updating</i> è uno, ma non l'unico, dei possibili modi in cui si fa procedere il Discorso)	Regola che garantisce la coerenza del discorso, prescrivendo che l'aggiornamento (non solo quello <i>topic-based</i>) avvenga attaccando ogni nuovo segmento di Discorso ad uno precedente	Esempi costruiti di testi monologici composti da frasi ben formate	Implementati vari algoritmi DRT in LISP e PROLOG	?
Prévot 2001	Semantica del Discorso orientamento descrittivo	SDRT estesa ai dialoghi come in Asher & Lascarides (1998), ulteriormente integrata con <i>Topic Structure</i> (ispirata a van Kuppevelt) che si sviluppa parallelamente alla <i>Segmented Discourse Representation Structure</i>	Inteso come in van Kuppevelt, come una (domanda) subtopic	Non indagata	Inteso come il <i>Discourse Topic</i> di van Kuppevelt, chiamato <i>Global Topic</i> o <i>Dialogue Topic</i> , e identificato nel corpus considerato con una <i>route explanation</i>	Strutturante, sia attraverso la creazione di strutture gerarchiche domanda-risposta, sia attraverso l'aggiornamento della SDRS	∅	Dialoghi <i>ruote-explanation</i> (un caso specifico di dialoghi <i>query based</i> e <i>task-oriented</i>)	∅	Realizzata su un corpus di 15 dialoghi <i>ruote explanation</i> in francese parlato semi-spontaneo

Tabella comparativa: I modelli computazionali

AUTORE	ORIENTAMENTO	FRAMEWORK	NOZIONE DI TOPIC (DI FRASE O ENUNCIATO)	INTERPRETAZIONE SEMANTICA DEL TOPIC	NOZIONE DI DISCOURSE TOPIC	FUNZIONE DEL TOPIC NEL DISCORSO	NOZIONE DI RELEVANCE	TIPOLOGIA TESTUALE CONSIDERATA	IMPLEMENTAZIONI	TEST SU CORPORA
<p>1) Grosz <i>et al.</i> 1983, 1995 + 2) Grosz & Sidner 1986</p> <p>Le due proposte sono complementari</p>	<p>Analisi del Discorso; orientamento computazionale-teorico</p>	<p>1) Modello per la risoluzione della referenza nel discorso, attraverso un meccanismo di <i>Centering</i> 2) Modello della Struttura del Discorso (composto da: struttura linguistica, struttura intenzionale, struttura attenzionale + un meccanismo di <i>Focusing</i>)</p>	<p>Definito come <i>Center</i> dell'enunciato (modello 1)</p>	<p>Un oggetto semantico, che denota un'entità che è oggetto di referenza</p>	<p>Definito come <i>focus space</i>, segmento di discorso (composto da uno o più enunciati) caratterizzato da un'intenzione e che comprende anche oggetti, proprietà, relazioni (modello 2)</p>	<p>1) Il <i>Centering</i> risolve la referenza tra gli enunciati del Discorso, garantendo la coerenza locale; 2) Il <i>Focusing</i>, cooperando con l'intenzionalità e la struttura linguistica, struttura il Discorso in segmenti (di uno o più enunciati), garantendo la coerenza globale</p>	<p>Può essere intesa come coerenza, globale (tra segmenti del discorso) e locale (tra gli enunciati in un segmento del Discorso)</p>	<p>1) Enunciati costruiti e ben formati 2) Testi monologici espositivo-retorici; dialoghi task-oriented</p>	<p>1) Realizzati numerosi algoritmi per il <i>Centering</i> (cfr. Miltakaki 2002) 2) Grosz & Sidner propongono un modello computabile con una struttura a pila (<i>Focus stack</i>) che ha trovato implementazioni concrete in vari sistemi (ad esempio TEAM e XTRA, cfr. Bos 1995: 185)</p>	<p>?</p>
<p>Hearst 1997</p>	<p>Analisi del Discorso; rientamento computazionale applicato</p>	<p>Modello della struttura del Discorso a livello di organizzazione in paragrafi</p>	<p>∅</p>	<p>∅</p>	<p>Chiamato <i>subtopic</i>, inteso come l'argomento che caratterizza una unità multi-paragrafo (un segmento di testo) attraverso pattern di cooccorrenza lessicale; distinto dal <i>main topic</i>, l'argomento generale di cui tratta il testo</p>	<p>Strutturante, attraverso un meccanismo di TextTiling, cioè individuazione di unità multi-paragrafo di testo</p>	<p>∅</p>	<p>Testi espositivi</p>	<p>Implementato un algoritmo per il TextTiling che individua i segmenti sub-Topic identificando i Topic shift attraverso pattern di cooccorrenza e distribuzione lessicale</p>	<p>?</p>

