

Cristiano Chesi, Andrea Moro

Il divario (apparente) tra gerarchia e tempo

(doi: 10.1422/89315)

Sistemi intelligenti (ISSN 1120-9550)

Fascicolo 1, Aprile 2018

Ente di afferenza:

Università degli Studi di Pavia (Unipv)

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.

Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

L'articolo è messo a disposizione dell'utente in licenza per uso esclusivamente privato e personale, senza scopo di lucro e senza fini direttamente o indirettamente commerciali. Salvo quanto espressamente previsto dalla licenza d'uso Rivisteweb, è fatto divieto di riprodurre, trasmettere, distribuire o altrimenti utilizzare l'articolo, per qualsiasi scopo o fine. Tutti i diritti sono riservati.

IL DIVARIO (APPARENTE) TRA GERARCHIA E TEMPO

1. INTRODUZIONE

Questo lavoro si pone l'obiettivo di dimostrare come per spiegare alcuni semplici fenomeni linguistici, una certa concezione di tempo sia da un lato fuorviante, se utilizzata indipendentemente, dall'altro cruciale, se riconciliata con una rappresentazione gerarchica delle espressioni linguistiche.

Il caso in esame riguarda quella facoltà, squisitamente linguistica, che permette di potersi «auto-riferire» allo stesso argomento all'interno di un enunciato, ad esempio, utilizzando un pronome clitico riflessivo come nella frase «*Adamo si dispera*». L'«auto-riferimento», o più precisamente la *co-referenza* tra pronome clitico «*si*» e il sintagma nominale «*Adamo*», sembra banale da risolvere nella frase precedente. In realtà, qua mostreremo come l'apparente semplicità di questa costruzione celi diverse insidie che ci porteranno a riflettere, in primo luogo, sulla natura della rappresentazione linguistica mentale necessaria per descrivere correttamente ogni rilevante declinazione empirica del fenomeno, in secondo luogo, sulla portata cognitiva dei concetti di «*aspettativa*» e «*asimmetria*» indotte dal tempo (inteso come sequenza di eventi discreti) in cui gli elementi vengono processati.

In pratica, cercheremo di mostrare come la semplice posizione (e co-distribuzione in termini puramente statistici) dei potenziali referenti (e.g. «*la compagna di Adamo si dispera*» vs. «*Adamo si dispera*» vs. «il fatto che *si* sia lasciato convincere da Eva, preoccupa molto *Adamo*») sia fuorviante nel cercare un principio di risoluzione di questa anafora pronominale che possa competere con un'analisi strutturale gerarchica (quale quella che classicamente si è soliti adottare da Chomsky 1981; 1986 in poi).

Nonostante l'evidente superiorità esplicativa di un approccio che basa la propria analisi su una rappresentazione mentale puramente «gerarchica» tra gli elementi in gioco, mostreremo come l'asimmetria prodotta dal flusso delle parole rappresenti un fondamentale indizio da sfruttare nell'analisi della co-referenza anaforica, permettendoci così di spiegare

elegantemente alcuni sottili aspetti che resterebbero altrimenti oscuri. Vedremo anche come questa prospettiva ci permetta di generalizzare il concetto di «aspettativa» e «presupposizione» in modo da risultare cognitivamente (e non solo linguisticamente) interessante.

Inizieremo il nostro percorso introducendo i concetti di *tempo* e *gerarchia* (§ 2) in modo funzionale al dato linguistico che intendiamo discutere (§ 3), sottolineando quelli che sono i principali approcci classici al problema sia da un punto di vista puramente strutturale (§ 3.1) che distribuzionale (ad esempio discutendo l'approccio al linguaggio con Reti Neurali Ricorrenti Semplici, § 3.2). Mostriamo quindi come il primo approccio, sebbene superiore in termini esplicativi (§ 4.1, § 4.2) necessiti di un'integrazione del concetto di *tempo* (§ 5) per poter catturare alcune costruzioni problematiche. In conclusione, rifletteremo sulla portata cognitiva più ampia (rispetto al ristretto dominio linguistico che ha originato questa riflessione) che gli strumenti qua introdotti potrebbero avere (§ 6).

2. BACKGROUND TEORICO: INTRODUZIONE AL TEMPO E AL CONCETTO DI GERARCHIA

Iniziamo considerando il *tempo* come un ordine totale tra eventi discreti. Se ci concentriamo sui fenomeni linguistici, un siffatto ordine totale è sempre stabilito tra precisi eventi discreti, cioè la pronuncia/percezione delle parole (o dei morfemi che le compongono).

Va sottolineato subito che questa non è una condizione necessaria: se si pensa a questo ordine totale come a qualcosa di imposto dalla modalità fono-articolatoria (che solitamente viene usata per veicolare contenuti linguistici), potremmo ipotizzare, pensando alle lingue dei segni, che utilizzando una modalità visivo-gestuale, esse potrebbero prescindere da un ordine lineare totale tra gli elementi all'interno degli enunciati prodotti: in pratica, sfruttando la possibilità di co-articolazione di più segni in contemporanea nello spazio segnico, le lingue dei segni potrebbero realizzare due o tre argomenti di un preciso predicato esattamente nello stesso istante. Ad esempio, prendiamo una frase come (1):

(1) Eva mangia la mela

L'ordine delle parole è precisamente (ed uniformemente) determinato sia quando produciamo che quando recepiamo la frase. Sebbene ci siano ordini alternativi possibili, è evidente a qualsiasi parlante nativo come solo pochissime permutazioni siano accettabili in Italiano:

(2) a. LA MELA mangia Eva!
b. *Eva la mangia mela

Ad esempio (2.a), con la giusta prosodia (cioè con una focalizzazione che porta l'accento sull'informazione nuova/contrastiva, ovvero il sintagma nominale in maiuscoletto, «LA MELA»), esprime esattamente la stessa azione e gli stessi argomenti espressi da (1) (con prosodia neutra invece si dovrebbe intendere che è la mela a mangiare Eva!), mentre (2.b) è chiaramente un ordine impossibile.

Ai fini di questo articolo è fondamentale soffermarsi sul fatto che, qualunque siano gli ordini ammissibili, in nessun linguaggio verbale è possibile co-articolare due elementi, ad esempio i due argomenti del predicato «mangiare» contemporaneamente come rappresentato nell'esempio (3) (le parentesi graffe indicano lo stesso istante di co-articolazione e il diacritico «|» separa i due elementi logici da co-articolare):

(3) *{Eva | la mela} mangia

Per quanto ne sappiamo, nessuna lingua dei segni, così come nessuna lingua verbale, permette una simile perfetta co-articolazione di elementi lessicali senza che tra di essi sia introdotta una chiara asimmetria temporale: ad esempio nell'American Sign Language (Napoli e Sutton-Spence 2010), in una frase che può essere tradotta come «un uomo sta cercando il gatto che miagola sull'albero di Paul Scott» (glossata come «A man looking for the meowing cat in Paul Scott's "Tree"») gli argomenti (*un uomo*, *il gatto* e *l'albero di Paul*) vengo co-articolati nello spazio in un tempo determinato, ma solo se si introduce un'asimmetria e una precisa sequenza di articolazione dei tre oggetti la frase ha un senso: anche se l'argomento *un uomo* viene rappresentato dal dito della mano destra alzato, mentre il terreno dove si trova *l'albero* è segnato contemporaneamente con la mano sinistra piatta, solo il movimento e la direzione degli occhi del parlante (lo guardo dell'uomo), unitamente ai tempi del movimento della bocca (il gatto che miagola) pongono una chiara e necessaria asimmetria temporale che permette di capire esattamente chi fa cosa, dove e quando.

Proviamo ad assumere quindi che il «tempo» sia davvero un elemento essenziale di ogni linguaggio umano, o, più precisamente, che il tempo introduca una relazione asimmetrica (totale) negli eventi linguistici tale da risultare cruciale nell'interpretazione dell'enunciato.

In pratica, per puro gusto speculativo, anche se avessimo a disposizione una modalità di comunicazione telepatica, questa assunzione «temporale» ci porterebbe a ipotizzare che una sequenzializzazione del messaggio da trasmettere telepaticamente sarebbe comunque necessaria.

Se di primo acchito potrebbe sembrare scontato che le lingue verbali (così come quelle segniche) debbano necessariamente sviluppare il loro flusso in modo temporalmente definito e che un prima e un dopo sia sempre presente sia in fase di comprensione che di produzione di un enunciato, va però ricordato che in quasi nessun modello computazionale

(ad eccezione dei modelli di *parsing*) e formale, questo vincolo fondamentale viene preso in considerazione in modo sostanziale. L'ordine lineare, in effetti, da un lato, dal punto di vista formale, viene considerato come un mero epifenomeno della modalità fonico-articolatoria (e potenzialmente di quella visivo-gestuale) totalmente irrilevante per la descrizione grammaticale (Chomsky 1995; 2008), dall'altro lato, quello della performance linguistica, viene considerato come l'unica fondamentale fonte di informazioni di cui un apprendente debba aver bisogno per acquisire le fondamentali proprietà della lingua (Tomasello 2003). L'idea presentata in questo articolo sottolinea che queste due posizioni, potenzialmente inconciliabili, in realtà debbano essere, almeno parzialmente, entrambe considerate per migliorare il modello di descrizione grammaticale sia empiricamente che cognitivamente. Iniziamo quindi fornendo alcuni dettagli in più sulle due posizioni.

2.1. *L'approccio gerarchico e l'epifenomenalità dell'ordine lineare*

Seguendo la prospettiva formale, proviamo ad adottare un approccio Minimalista (parafrasando la massima di Occam: «solo ciò che è strettamente necessario verrà incluso nella teoria linguistica»). Secondo Chomsky (1995), dato un lessico dotato di morfemi e parole con associati tratti fonetici e semantici utili rispettivamente alla pronuncia e all'interpretazione delle parole, solo un'operazione di base di combinazione di questi elementi lessicali è necessaria per la creazione della struttura frasale fondamentale per interpretare correttamente la frase: l'operazione in questione si chiama *Merge* e permette di formare insiemi («costituenti») gerarchicamente sempre più grandi. Il risultato di *Merge* (A, B), dove A e B sono due elementi lessicali qualsiasi, sarà quindi uguale a [A, B], cioè un insieme non ordinato formato dall'elemento A e dall'elemento B. Tale insieme potrà essere oggetto di ulteriori operazioni di *Merge*: *Merge* (C, [A, B]) = [C [A B]] e così via. Paradossalmente, seguendo questa strada che formalmente e computazionalmente è perfettamente sensata, si verrebbe a costruire la frase in (1) da destra a sinistra utilizzando (almeno) tre operazioni di *Merge* esattamente in quest'ordine:

- (1) i. [la mela]
- ii. [mangia [la mela]]
- iii. [Eva [mangia [la mela]]]

Sebbene la reale *derivazione* della frase sia qua semplificata di molto, la struttura del problema non viene minimamente snaturata: la teoria minimalista prevede che la frase venga *generata* seguendo un ordine totalmente indipendente da quello con cui viene *processata*. L'ordine in cui poi le parole vengono pronunciate e percepite, secondo la declinazione più accreditata della teoria, in realtà è imposto dal sistema

motorio-percettivo (noto come *Phonetic Form, PF*), esterno al motore generativo linguistico e a cui il motore generativo deve interfacciarsi così come dovrà interfacciarsi con il sistema interpretativo (anche detto *Logic Form, LF*). Fox e Pesetsky (2005) parlano esplicitamente di un «Phonetic Form effect» riferendosi alla necessità, esclusivamente imputabile alla modalità fonico-articolatoria, di linearizzare una struttura gerarchica prodotta da una sequenza di operazioni di Merge con il fine di renderla pronunciabile. La grammatica, secondo l'approccio formale, per computare le relazioni rilevanti all'interno di una frase, ha solo bisogno della gerarchia, cioè delle relazioni di inclusione espresse dalle parentesi quadre nell'esempio sopra riportato ed in questo senso [Eva [mangia [la mela]]] è esattamente equivalente a [[[la mela] mangia] Eva]), salvo poi doversi interfacciare con PF e quindi decidere se la prima o la seconda versione dovranno essere realizzate. Nel prossimo capitolo approfondiremo adeguatamente questo aspetto, adesso introduciamo l'altro punto di vista sulla frase in (1) in opposizione a (2.a) o (2.b).

2.2. *L'approccio usage-based e la crucialità dell'ordine lineare*

Secondo il modello che privilegia l'ordine delle parole così come vengono prodotte e recepite, la frequenza d'uso con cui la co-occorrenza di certi termini si realizza è fondamentale per spiegare la nostra abilità linguistica e prevedere la difficoltà che avremo con particolari tipi di frasi. Tale approccio, «basato sull'uso» («usage-based», Tomasello 2003) per cui è rilevante la frequenza con cui una costruzione si presenta al parlante, permette di spiegare alcuni fenomeni interessanti. Per fare un esempio, prendiamo in esame un corpus dell'italiano di ragionevole grandezza (ad esempio l'insieme delle pagine Web scritte in Italiano). All'interno di questa collezione, potremo trovare un buon numero di co-occorrenze di parole che sono co-occorrenti anche nelle frasi (1) e (2.a), mentre ci sono elementi co-occorrenti in (2.b) che non saranno assolutamente frequenti nello stesso corpus. Se in effetti verifichiamo questa ipotesi, semplicemente tentando delle «google-battles» (confronto del conteggio del numero di documenti restituiti in risposta ad una query i cui termini vengono messi tra virgolette in modo da garantire che solo le sequenze esatte delle parole in questione vengano ricercate) tra i vari tri-grammi che compongono la frase (sequenze sovrapposte di 3 parole), questo sarà il risultato nei tre casi in esame:

- (1) «eva mangia la» 1.700 occorrenze + «mangia la mela» 70.600 occorrenze = 72.300 risultati
- (2) a. «la mela mangia» 466 occorrenze + «mela mangia eva» 0 risultati = 466 risultati
- b. «eva la mangia» 7 occorrenze + «la mangia mela» 0 risultati = 7 risultati

Dal numero di occorrenze totale trovate, si può dedurre che i trigrammi che compongono la frase in (1) sono molto comuni in Italiano (almeno secondo Google), quelli che compongono la seconda frase in (2.a) lo sono molto meno e quelli che compongono la frase in (3) sono rarissimi. Da ciò si può dedurre che (1) è perfettamente grammaticale e facile da processare, (2.a) può essere grammaticale ma decisamente più «rara» di (1) e (2.b) è probabilmente agrammaticale. L'ordine lineare (da cui parte la scomposizione in n-grammi), da questa prospettiva, ci permetterebbe di dedurre sia la grammaticalità che la facilità di processamento.

2.3. *Alcuni problemi*

Il dibattito tra le due prospettive precedentemente discusse, quella basata sulla gerarchia (talvolta genericamente, e forse impropriamente, definita «rule-based», § 2.1) e quella basata sulla frequenza d'uso («usage-based», § 2.2) è a lungo stato aspro ed è tuttora irrisolto: da un lato si sono collezionate evidenze palesi che dimostrano come, ad esempio, dalla semplice analisi della distribuzione degli ausiliari non si sia in grado di inferire alcune proprietà delle costruzioni interrogative in lingue quali l'Inglese o il Norvegese che in realtà un bambino è in grado di applicare molto precocemente (e.g. invertire correttamente l'ausiliare appropriato con il soggetto della frase principale tenendo conto dei vincoli delle costruzioni a verbo secondo, Westergaard 2009), dall'altro si è cercato di mostrare come addirittura il cavallo di battaglia dell'approccio «rule-based», cioè la ricorsività nelle strutture sintattiche, e la difficoltà di processamento in precisi contesti ricorsivi (i.e. incasamento centrale), classicamente considerati come casi di limitazioni della «performance» linguistica (Chomsky 1957), siano in realtà più adeguatamente e semplicemente descrivibili attraverso l'interazione tra esperienza linguistica e vincoli del meccanismo di apprendimento generalizzato (Christiansen, MacDonald 2009).

Approcciandosi però criticamente all'idea che la semplice co-distribuzione degli elementi sia fondamentale per determinare la facilità di processamento e dedurre la grammaticalità di un enunciato, non è difficile mostrare come in verità la cruda frequenza di co-occorrenza degli elementi all'interno di una frase possa risultare fuorviante sia nel sussumere il concetto di grammaticalità, sia nel dedurre dove si presenteranno le difficoltà di processamento:

- (3) a. *Eva porta la apre
- b. LA PORTA apre Eva (stessa intonazione di (2.a))

In (3.a) avremmo «Eva porta la» che registra 3.170 occorrenze e «porta la apre» che ne registra ben 23.100. Da ciò si dovrebbe dedurre, come abbiamo fatto in § 2.2, che la frase (3.a) dovrebbe essere perfettamente grammaticale, contraddicendo l'intuizione di qualsiasi parlante nativo. (3.b), d'altro canto, registrando 4.230 occorrenze con «la porta apre» e nessuna occorrenza con «porta apre Eva» farebbe pensare che la sequenza «porta apre Eva» è non solo strana ma addirittura agrammaticale, contrariamente ai fatti. Declinazioni più o meno sofisticate di questa intuizione hanno permesso di sviluppare modelli comunque interessanti che permettono, ad esempio, di prevedere con una ragionevole accuratezza la classe grammaticale e altri tratti morfo-sintattici della parola che seguirà una sequenza di altre parole (Elman 1990) e addirittura di simulare un comportamento «umano» nel processamento di strutture gerarchiche molto complesse (incassamento centrale vs. incassamento cross-seriale, Christiansen e Chater 1999) come ci accingeremo a mostrare in § 3.2.

Va notato, inoltre, che anche l'approccio gerarchico potenzialmente potrebbe presentare alcuni problemi: se l'operazione di Merge fosse davvero insensibile all'ordine lineare (indipendentemente da requisiti imposti da PF), [*Adamo* [*si dispera*]] e [[*si dispera*] *Adamo*] dovrebbero essere ugualmente plausibili in termini di linearizzazione e la co-referenza del riflessivo con il sintagma nominale *Adamo* dovrebbe uniformemente essere soddisfatta nel primo come nel secondo caso senza nessun'altra assunzione strutturale. In realtà, in lingue come l'Italiano, la seconda opzione di linearizzazione «*si dispera Adamo*» è molto più marginale della prima e legittima, come nell'esempio (2.a), solo in particolari condizioni di focalizzazione: «*si dispera*» deve essere focalizzato come nuova e/o inattesa informazione, legittimando così l'ordine verbo-soggetto. Tale ordine è legittimo in molti contesti in Italiano (ad esempio con un verbo inaccusativo: «cade la mela») che però non richiedono assolutamente le stringenti condizioni «pragmatiche» che invece sembrano necessarie nell'interpretazione dell'anafora riflessiva. L'approccio gerarchico ha quindi due possibilità: il legamento dell'anafora pronominale non riguarda solo una configurazione strutturale, ma sottostà a requisiti pragmatici esterni alla sintassi; l'idea che l'operazione di Merge sia sufficiente a spiegare ogni asimmetria strutturale necessaria va rivista e, potenzialmente, un'altra dimensione strutturale (quella temporale, appunto) dovrebbe essere inclusa nella nostra riflessione. In questo articolo perseguiremo questa seconda ipotesi.

3. UN DATO LINGUISTICO DA SPIEGARE: TEMPO E GERARCHIA NELL'ANAFORA PRONOMINALE

3.1. *La distribuzione dei pronomi clitici riflessivi*

In Italiano, come in molte altre lingue, esiste la possibilità, con certi tipi di predicati, di far «assorbire» al soggetto anche il ruolo di complemento oggetto. Ad esempio, in una frase come (4.a), il predicato «lavare» prende due argomenti, un agente «Adamo» e un paziente «Eva». Nella sua variante (4.b) il ruolo di «paziente» viene assorbito dal soggetto del predicato attraverso l'uso del pronome riflessivo «si», appunto co-referente con il soggetto (si vedano gli indici identici). Un'equivalente semantico di (4.b) potrebbe essere (4.c), ma non (4.d) o (4.e), in cui «lui» e la seconda occorrenza di «Adamo», rispettivamente, non possono essere co-referenti con il soggetto «Adamo»:

- (4) a. Adamo lava Eva
- b. Adamo_i si_i lava
- c. Adamo_i lava [se stesso]_i
- d. Adamo_i lava lui_{*i/j}
- e. Adamo_i lava Adamo_{*i/j}

Queste restrizioni, vengono spiegate dalla «teoria del legamento» (Chomsky 1981), una teoria cioè che individua le restrizioni strutturali che vincolano certe possibilità di co-referenza. In concreto, il paradigma in (4.b/c), (4.d) e (4.e) può essere rispettivamente descritto dai tre principi che seguono:

- (5) Principi del Legamento
 - Principio A* – un'anafora deve essere legata nel suo dominio locale
 - Principio B* – un pronome deve essere libero nel suo dominio locale
 - Principio C* – un'espressione referenziale è sempre libera

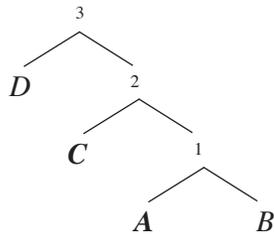
In pratica, il principio che regola il legamento del pronome riflessivo è il principio A. Per approssimare il concetto di «dominio locale», dobbiamo ricorrere a una relazione inerentemente gerarchica, ovvero il C-Comando:

- (6) C-Comando
 - α C-comanda β , se il primo *nodo* che domina α , domina anche β .

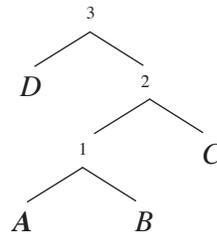
«Locale» viene invece approssimato al fatto che il dominio di C-comando debba essere «il più ristretto possibile» ovvero, semplificando l'idea espressa in Chomsky 1986, che il sintagma nominale co-referente

all'anafora in questione debba trovarsi, al massimo, all'altezza del primo argomento verbale (cioè la posizione del soggetto). Per mostrare in linea teorica come questo principio operi, esaminiamo una serie di strutture gerarchiche possibili quali quelle in (7) (pensiamole tutte generate dal basso verso l'alto attraverso l'applicazione di successive operazioni, numerate, di Merge). Osserviamo che sia in (7.a) che in (7.b) l'elemento *C* comanda l'elemento *A* poiché il primo «nodo» che domina *C*, cioè 2, domina anche *A* in entrambi i casi (la relazione di dominanza è transitiva, cioè se 2 domina *C* e *C* domina *A* allora anche 2 domina *A*). Questo ci mostra come la nozione di *C*-Comando sia insensibile all'ordine lineare (normalmente si considera una relazione lineare un ordine totale tra le foglie dell'albero, nel nostro caso, in (7.a) avremo quest'ordine totale «*D, C, A, B*» e in (7.b) quest'altro «*D, A, B, C*»):

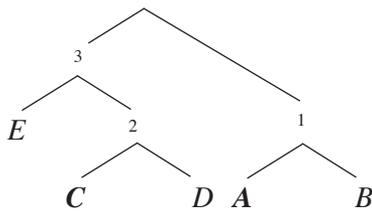
(7) a. [₃ *D* [₂ *C* [₁ *A B*]]]



b. [₃ *A* [₂ [₁ *C D*] *B*]]]



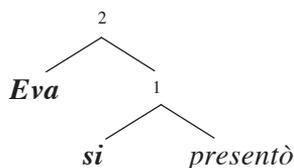
c. [₄ [₃ *E* [₂ *C D*]] [₁ *A B*]]]



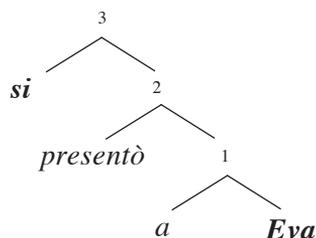
La stessa relazione di *C*-comando non è presente in (7.c) visto che il primo nodo che domina *C*, cioè 2, non domina *A*, quindi *C* non *C*-Comanda *A* in (7.c).

Questi alberi sono particolarmente interessanti perché spiegano esattamente il problema rappresentato dagli esempi in (8), per cui viene indicata la struttura (semplificata):

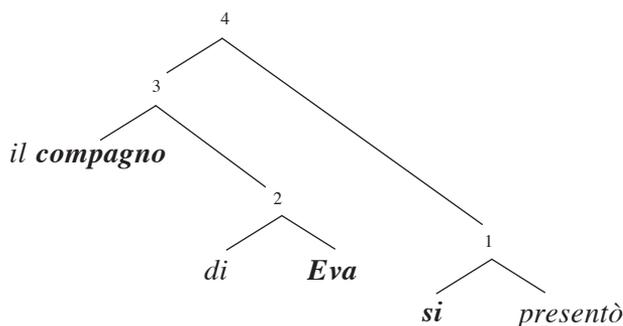
(8) a. Eva si presentò



b. Si presentò a Eva



c. Il compagno di Eva si presentò



In (8.a), Eva C-comanda il clitico riflessivo e quindi, secondo il principio A della Teoria del Legamento in (5), è un candidato ideale come referente. Questo non avviene né in (8.b), dove il primo nodo che domina «Eva», cioè 1, non domina il pronome riflessivo «si», né in (8.c), dove di nuovo, il nodo 2, che domina «Eva», non domina «si» e questo esclude (correttamente) che il referente giusto dell'anafora pronominale sia «Eva».

Vari studi mostrano come il Principio A della teoria del legamento, in particolare in lingue come l'italiano in cui il pronome clitico è coinvolto (McKee 1992), oltre alla variante pronominale piena («se stesso/a»), sia dominato completamente dai bambini dai 3 ai 5 anni.

Enunciamo a questo punto ancora due casi che rendono la faccenda alquanto intrigante da un punto di vista cognitivo: nel primo caso, vediamo come il Principio A in (5) non sia sufficiente a spiegare quali dei due referenti, entrambi in posizione di C-Comando rispetto al riflessivo, sia il co-referente giusto (9); nel secondo caso, (10), vedremo come l'ordine lineare sia insufficiente a determinare la relazione di co-referenza necessaria (due nomi coordinati, peraltro in posizione post-verbale fanno entrambi da co-referente al riflessivo):

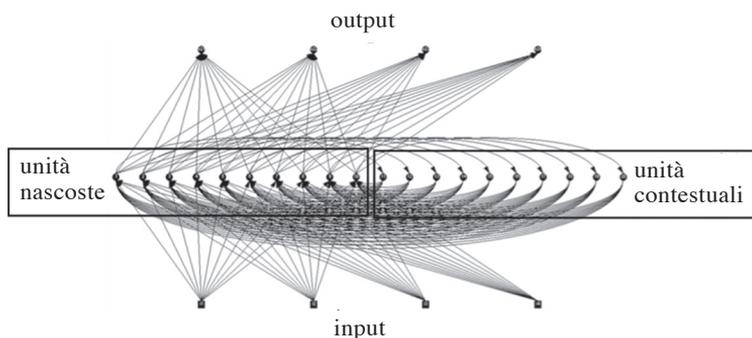
(9) E così Eva_i, ADAMO_j si_{i/*j} fece presentare!

(10) Su questa idea si_i confrontano [Adamo ed Eva]_j

3.2. *La creazione delle aspettative basate sulla sequenza delle osservazioni*

Mettiamo per un attimo da parte le considerazioni sulla rilevanza della rappresentazione gerarchica degli enunciati e concentriamoci su ciò che di interessante ci dice la sequenza in cui le parole vengono prodotte e lette/ascoltate. Elman (1990; 1991; 1993) ha tentato di rappresentare il flusso temporale che caratterizza il processamento linguistico utilizzando delle reti neurali artificiali ricorrenti semplici (Simple Recurrent Networks SRN) come quella rappresentata in (11):

(11) Reti Ricorrenti Semplici (SRN)



Nelle SRN una copia delle unità nascoste (hidden layer) viene mantenuta «in memoria» al tempo t (le unità contestuali mantengono una copia uno a uno del livello di attivazione delle unità presenti nell'hidden layer al tempo t) e data in pasto alla rete al tempo $t+1$ insieme al nuovo input che segna una nuova epoca di processamento della rete. Reti ricorrenti di questo tipo possono essere valutate rispetto alla loro abilità di «prevedere la parola successiva» basandosi sulle frasi precedentemente mostrate loro. In pratica, data una frase come quella in (1) («Eva mangia la mela»), la rete, quando riceve in input «Eva» dovrà rispondere «mangia», quando riceve in input «mangia» dovrà rispondere «la» e quando riceve «la» dovrebbe rispondere «mela». Tra le molte possibili previsioni che la rete potrà fare essendo stata esposta a moltissime (ripetizioni di) frasi dell'Italiano, la rete finirà per prevedere non tanto la parola esatta, ma la categoria e possibilmente diversi tratti morfologici della parola che verrà dopo. La qualità della rappresentazione del problema viene valutata effettuando una Principle Component Analysis (PCA, Elman

1991) rispetto all'attivazione delle unità del layer nascosto. Se due parole «simili», cioè con categoria uguale e tratti morfologici identici (tipo numero, persona, genere, animatezza ecc.) producono un'attivazione simile nel layer nascosto, allora la rete avrà in qualche modo astratto e appreso tale regolarità morfo-sintattiche e semantiche; quindi se dopo un soggetto singolare la rete prevedrà un verbo singolare, potremmo inferire che la rete ha acquisito la corretta dipendenza tra i cluster di parole del soggetto e quelle del verbo. Va sottolineato che non solo le dipendenze «locali» sembrano essere apprese dalla rete (relazioni quali appunto quella che può intercorrere tra la posizione del soggetto e quella del verbo quando questi due elementi si trovano contigualmente l'uno accanto all'altro), ma anche in un contesti di relazioni non locali (ad esempio quando il soggetto e il verbo sono separati da una frase relativa come nel caso seguente: «*Eva*, che noi conosciamo molto bene, *mangia* la mela»; se il verbo «mangia» viene previsto con maggiore probabilità rispetto a «mangiamo» la rete avrà «riconosciuto» che il soggetto corretto è quello più distante, «Eva», e non quello più vicino, cioè «noi»).

All'approccio di Elman va riconosciuto il merito di aver formalizzato il concetto di «tempo» esattamente come lo abbiamo definito nel primo paragrafo. Ciò, stando agli interessanti risultati mostrati dagli esperimenti di Elman, pare essere sufficiente a catturare molte regolarità legate alle aspettative sulla parola seguente all'interno di una frase che viene percepita e/o prodotta.

Le SRN sembrano inoltre in grado di simulare proprietà ancora più sottili della performance umana quando sotto analisi finiscono le strutture ricorsive, cioè, in termini più precisi, quando una struttura frasale di un certo tipo contiene al suo interno una struttura frasale esattamente dello stesso tipo (ad esempio quando un soggetto di una dichiarativa viene modificato da una frase relativa e così via: «il cane che il gatto che il topo rincorse scappò morse»). Christiansen e Charter (1999) mostrano come una rete ricorrente come quella di Elman, con un numero sufficiente di nodi nel layer nascosto, sono in grado di apprendere tre dipendenze non locali sia di tipo «incassato» (e.g. $a b c c b a$) che di tipo cross-seriale ($a b c a b c$), peraltro comportandosi esattamente come un essere umano nei due casi, cioè trovando più facile il secondo tipo di dipendenze rispetto al primo, cosa non scontata stando alla Gerarchia di Chomsky, visto che, secondo tale gerarchia, grammatiche con la potenza generativa delle grammatiche Context-Free (Context-Free Grammars, CFGs), sono in grado di catturare le dipendenze di tipo incassato, ma non quelle di tipo cross-seriale. Questo aspetto è stato discusso approfonditamente, sia da un punto di vista psicolinguistico che computazionale e neuro-linguistico (Chesi e Moro 2014).

L'unico modo per verificare se questa intuizione è sufficiente è metterla alla prova su specifici problemi empirici. Il fenomeno che abbiamo scelto come banco di prova è piuttosto semplice ed è appunto quello

della co-referenza con i riflessivi clitici in Italiano. Se osserveremo che l'ordine lineare o la struttura gerarchica, indipendentemente, riescono entrambi a render conto di ogni fenomeno osservato, il nostro lavoro sarà comunque servito a dimostrare l'equivalenza dei due punti di vista. Se, al contrario, uno dei due approcci risulterà superiore in termini esplicativi, dovremmo logicamente abbandonare l'altro. Se infine nessuno dei due approcci si mostrerà completamente adeguato, cercheremo di prendere una posizione sul tipo di revisione o integrazione necessaria per l'uno o l'altro approccio in modo da migliorarne l'adeguatezza esplicativa.

4. ASPETTI CRITICI DELL'UNA E DELL'ALTRA ANALISI

4.1. *Un semplice test per mostrare come le frequenze distribuzionali non siano sufficienti per inferire il legamento dei pronomi riflessivi*

Se le frequenze distribuzionali fossero sufficienti, da sole, ad apprendere i contesti di legamento dei pronomi riflessivi in Italiano, ci aspetteremmo di trovare un'evidenza numerica dei vari contesti sopra discussi. In particolare ci aspetteremmo che contesti in cui la coreferenza è possibile siano equamente distribuiti, affinché non subentri un fenomeno di iper-generalizzazione nell'assegnazione del referente, oppure, nel caso si riveli una forte asimmetria nella distribuzione dei vari contesti di co-referenza, dovremmo prevedere una proporzionale asimmetria a livello di difficoltà nel processamento delle diverse strutture così disegualmente distribuite. Ad esempio, stando a (8.a) («Eva_i si_i presentò») potremmo ipotizzare che il miglior referente è quello che occupa la posizione più vicina al riflessivo clitico. Ma questo è in contrasto con il dato in (8.c) («[Il compagno [di Eva_j]]_i si_{i/*j} presentò»). Quindi, visto che (8.c) è possibile, ci aspetteremmo di trovarne un'evidenza in termini distribuzionali. Per la nostra analisi abbiamo utilizzato il corpus di Repubblica (Baroni *et al.* 2004). Il corpus è costituito da una collezione di articoli pubblicati sul quotidiano «La Repubblica» dal 1985 al 2000. Attualmente la collezione consiste di 380M tokens annotati semi-automaticamente per Part-of-Speech seguendo lo standard TEI (<http://www.tei-c.org>; l'annotazione prevede una cinquantina di categorie ispirate alle linee guida EAGLES, Monachini 1996). Pur non essendo ideale come «reference corpus» per l'Italiano (poiché completamente basato su un'unica risorsa, il quotidiano «La Repubblica» appunto) è sicuramente questo uno dei corpora più utilizzati in ambito linguistico, anche rispetto a corpora meglio bilanciati ma più piccoli (e.g. COLFIS, Bertinetto *et al.* 2005) o più estesi ma più rumorosi (e.g. itWaC, Baroni, Bernardini, Ferraresi e Zanchetta 2009).

Il nostro obiettivo era osservare la distribuzione del pattern «NP *si*» e «NP PP *si*». Utilizzando il servizio on-line che permette di interrogare

liberamente il corpus (<http://dev.sslmit.unibo.it>) abbiamo utilizzato la query CQL «[POS = “NOUN/NPR”] [WORD = “si” & POS = “CLI”]» ottenendo così 674.057. Analizzandone (manualmente) 1.000 occorrenze random dal campione recuperato, questa è la distribuzione che si osserva:

i. Circa il 51% dei clitici riflessivi ha il suo referente nel sintagma nominale immediatamente precedente (e.g. «[la camera]_i si_i appresta a discutere»); nel 4% dei casi si tratta di due NP congiunti, e.g. «[liberali e socialdemocratici]_i si_i sono schierati contro»);

ii. Circa il 22% dei riflessivi deve trovare il suo referente in una posizione precedente, ma non locale (15% dei casi [NP [PP]], e.g. «[il vertice [di Dublino]]_i si_i è dimostrato...»; 4% dei casi [NP [PP [PP]]], e.g. «[la marcia [dell'ultimo [dell'anno]]]_i si_i è svolta ...», 3% circa dei casi [NP [RC]], e.g. «[i lavoratori [che hanno superato i 50 anni di età]]_i si_i dovrebbero dimettere»);

iii. Circa l'8% dei riflessivi è co-referente con un sintagma nominale post-verbale (e.g. «ieri mattina si_i è riunito [[il consiglio] dei ministri]_i»)

iv. Nel 2% dei casi il referente è assente nel contesto frasale e inferibile solo nel contesto del discorso (e.g. «Il giorno di San Silvestro si_i è allontanato ...»)

v. Nel 17% dei casi, l'uso del «si» è in realtà impersonale e non riflessivo (e.g. «In Italia si segna di meno»).

Le considerazioni che si possono fare sono di due tipi: da un lato si deve osservare che non esiste nessun particolare tipo di evidenza che dimostri la marginalità/difficoltà del pattern ii. o iii. rispetto ad i. Dall'altra è che in circa la metà dei casi in cui il pattern [NP si] si presenta, in realtà la co-referenza non è assolutamente possibile. La domanda logica è quindi come il legamento delle anafore pronominali si possa apprendere sulla base della semplice evidenza distribuzionale.

4.2. *Due conferme sperimentali sull'impossibilità di inferire la gerarchia dalla distribuzione statistica delle parole o dalla loro semplice sequenza*

Per rafforzare l'idea che la struttura logica dell'interpretazione dell'anafora pronominale non sia facilmente apprendibile sulla base delle regolarità di distribuzione statistica, abbiamo effettuato due esperimenti. Il primo utilizzando varie tecniche di Machine Learning, il secondo implementando una SRN sul modello di quella proposta da Elman.

4.2.1. *Machine Learning e la risoluzione dell'anafora riflessiva*

Nel primo esperimento abbiamo creato due tipologie di vettori, rappresentanti le frasi da interpretare, composti dalla sequenza delle parole che compongono il contesto (6 parole precedenti e 6 parole successive) in cui si trovava il clitico riflessivo e che comprendeva il sintagma nominale ad esso co-referente.

La prima tipologia di vettori includevano sia il contesto destro che quello sinistro e tentavano di classificare il dominio della relazione anaforica come locale (il sintagma nominale precedente adiacente era il co-referente giusto, caso i. discusso nel § 4.1), separato da uno più elementi (il sintagma nominale co-referente era separato dall'anafora pronominale da almeno un sintagma preposizionale, o frase relativa, caso ii.), co-referente da identificare nel sintagma nominale in posizione post-verbale (caso iii.), assente (il co-referente era sottointeso, caso iv.), o «si» non usato come riflessivo (caso v.). Nel secondo caso, solo l'intorno sinistro veniva incluso (6 parole che precedevano il «si», in ogni caso era presente il corretto co-referente laddove la frase originale lo includeva) e la classificazione escludeva i casi con co-referente post-verbale.

Per l'esperimento sono stati creati file in formato «arff» (Attribute-Relation File Format) secondo lo standard proposto dal software WEKA utilizzato per la creazione dei classificatori (Witten, Frank, Hall e Pal 2016).

Il dataset consisteva delle 1.000 occorrenze annotate e classificate recuperate dal corpus Repubblica (§ 4.1). Per la valutazione si è utilizzata una cross-fold validation con suddivisione del dataset in 10 sotto-gruppi (in pratica il 90% delle osservazioni veniva utilizzato per il training e il 10% per il test); i sotto-gruppi venivano fatti ruotare sistematicamente dal set di training a quello di test e i risultati riportati rappresentano la media della performance dei classificatori nei vari test.

Questi i risultati degli esperimenti di classificazione a seconda dei diversi algoritmi utilizzati per le due tipologie di vettori:

	[- 6 parole] <i>si</i> [+ 6 parole]	[- 6 parole] <i>si</i>
ZeroR	51.0%	53.5%
OneR	52.7%	55.2%
NaiveBayes	48.2%	57.3%
SMO	54.4%	59.6%

ZeroR è l'algoritmo più semplice che mostra la baseline del problema di classificazione per i due classificatori (in pratica, se non si dovessero utilizzare regole, ma si dovesse semplicemente assegnare sempre la classe più probabile, ovvero la co-referenza locale, l'accuratezza della classificazione si aggirerebbe intorno al 51-53.5%). Utilizzando una sola

regola, inferita tra tutti i possibili tratti lessicali presenti nel classificatore (algoritmo OneR), la performance non migliora di molto (52.7-55.2%). L'algoritmo NaiveBayes sfrutta invece il teorema di Bayes e assume che gli attributi all'interno del vettore (le parole contenute nella frase) sono tutti equamente (ed indipendentemente) importanti; l'accuratezza della classificazione ottenuta va dal 48.2 al 57.3%.

Per capire con che facilità vengono confusi i referenti, se ad esempio prendiamo la migliore performance nel classificatore più semplice (quello che include il referente nel solo contesto di sinistra e utilizza l'algoritmo SMO, Sequential Minimal Optimization per allenare un Support Vector Classifier) e studiarne la matrice di confusione.

(12) Matrice di confusione dell'algoritmo SMO applicato al classificatore con solo contesto sinistro (6 parole) per 7 classi:

a	b	c	d	e	f	g	< classificata come
258	13	0	1	0	0	10	a = locale
54	26	0	0	0	0	7	b = NP PP
16	5	2	0	0	0	2	c = NP PP PP
10	2	0	0	0	0	3	d = NP RC
0	1	0	0	0	0	1	e = NP PP RC
5	2	0	0	0	0	6	f = assente
68	7	0	0	0	0	28	g = nessun riflessivo

Si può notare come il classificatore tenda frequentemente ad associare il pronome riflessivo al sintagma nominale più locale precedente, o, in alternativa, a dedurre che il «si» non sia un vero riflessivo. In conclusione nessun algoritmo ci pare minimamente soddisfacente se applicato alla pura distribuzione delle parole nel contesto di occorrenza dell'anafora pronominale.

4.2.2. *Simple Recurrent Networks e l'assunzione temporale nella risoluzione dell'anafora riflessiva*

Nel secondo esperimento abbiamo investito maggiormente sul concetto di «sequenza» delle parole e abbiamo provato a replicare l'approccio di Elman: in pratica abbiamo costruito una rete Semplice Ricorrente (utilizzando TLearn, Elman 1993) con 1511 unità di input e di output che codificano, secondo una codifica localista ortogonale (un nodo, un item lessicale) il lessico estratto dagli esempi annotati. La rete è stata dotata di 100 unità nascoste (copiate in altrettante unità del context layer). Alla rete sono state somministrati 544 esempi annotati che includono solo casi di co-referenza con sintagmi nominali precedenti il clitico riflessivo, per un totale di 5179 items in sequenza mostrati, ciascuno,

500 volte. La rete doveva apprendere a prevedere la parola successiva, ed in particolare, giunta al «si» che concludeva la frase, prevedere se si trattava di un caso di co-referenza «locale» (caso i.), non locale (caso ii.), coreferente assente (caso iv.) o clitico non riflessivo (caso v.).

La performance della rete sembra leggermente migliore rispetto a quella dei classificatori precedentemente utilizzati e questo, brevemente, è il riassunto dei migliori risultati di accuratezza ottenuti nelle varie classi (il test è stato effettuato su 60 esempi esclusi dal training set la cui distribuzione per classe era proporzionalmente uniforme a quella del training set):

(13) Accuratezza della classificazione nelle 4 classi:

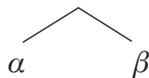
i. Co-referenza locale	92%
ii. Co-referenza non-locale	54%
iv. Co-referente assente	2%
v. Uso non riflessivo del «si»	9%

Come si può notare, anche in questo esperimento, la rete (come peraltro gran parte dei classificatori) ha una performance buona nel caso della co-referenza locale, ma decisamente mediocre negli altri casi e uniformemente ben al di sotto della performance ipotetica di qualsiasi parlante nativo dell'Italiano.

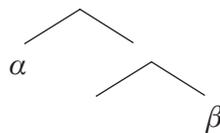
5. UN TENTATIVO DI RICONCILIAZIONE: FASI E ASPETTATIVE

Proviamo a questo punto a fare il punto delle osservazioni fin qui fatte nel tentativo di integrare la nozione gerarchica con una soluzione temporale che non banalizzi il problema come mostrato in § 4.2. Iniziamo sottolineando un tentativo interessante di conciliazione del concetto di linearizzazione con il concetto di gerarchia proposto attraverso la nozione di *antisimmetria* da Richard Kayne (1994). In breve, semplificando l'intuizione di Kayne, una struttura gerarchica non può essere simmetrica in termini di C-comando poiché altrimenti se un nodo α C-comandasse un nodo β e lo stesso nodo β C-comandasse α , allora non ci sarebbe modo di stabilire se α precede β o viceversa β precede α (14.a). Al contrario se α C-comandasse β e β non C-comandasse α , allora potremmo inferire che α precede β (14.b).

(14) a.



b.



Questa considerazione, qua puramente geometrica pone in realtà una fortissima restrizione sulle potenzialità delle descrizioni strutturali, escludendo a priori un notevole numero di possibilità senza ricorrere a «requisiti di interfaccia» à la Fox e Pesetsky. Mantenendo questa intuizione di fondo, proviamo quindi a riformulare la nozione di Merge in modo da renderla «antisimmetrica» (si vedano Phillips 1996 e Chesi 2012; 2015 per una trattazione formalmente più dettagliata di questi aspetti in termini di sequenza computazionale delle operazioni):

(15) Merge (antisimmetrico)

l'operazione base di costruzione della struttura si riduce sempre all'unione di un elemento incompleto (che include cioè un'aspettativa strutturale sul suo potenziale completamento) con uno completo (lessicale o prodotto di precedenti operazioni di Merge)

Se assumiamo che «l'elemento completo» possa essere un item lessicale o un costituente prodotto di diverse operazioni di Merge andate a buon fine, mentre l'elemento incompleto (atteso) è un costituente vuoto lessicalmente, specificato solo in termini di tratti da lessicalizzare obbligatoriamente, allora otteniamo subito una serie di conseguenze interessanti:

1. l'operazione di Merge sarà sempre un'operazione binaria che unificherà un (nuovo) item recuperato dal lessico con un'aspettativa strutturale precedentemente creata;

2. i nuovi item introdotti potranno generare nuove aspettative strutturali (e potenzialmente rivedere quelle vecchie, ma di questo non ci occuperemo qua);

3. se gli item introdotti non soddisfano tutte le aspettative in attesa, l'item resterà in «attesa di integrazione» e la frase sarà agrammaticale finché l'integrazione non sarà avvenuta.

Chiamiamo «fase» il dominio di un'aspettativa. Ipotizziamo che una fase sia istanziata dai tratti di selezione associati agli elementi lessicali: ad esempio un verbo come *mangiare* sarà dotato di due tratti di selezione per due sintagmi nominali distinti che creeranno l'aspettativa rispettivamente di un agente (e.g. *Eva*) e di un paziente (e.g. *la mela*).

Rispetto al nostro caso della co-referenza anaforica pronominale, adottiamo l'idea di Schlenker (2005) rivisitata ed arricchita da Bianchi (2009) per un'implementazione di una «memoria referenziale» associata ad ogni fase: ogni sintagma nominale della fase principale sarà «archiviato» nella memoria di questa fase come potenziale referente da utilizzare in contesti in cui un referente è necessario (come appunto in presenza di un clitico riflessivo). Più sarà recente il suo inserimento in memoria, più il referente risulterà saliente. Tale salienza, contrasta però con l'annidamento delle aspettative: l'idea di fase concepita come colle-

gata alla selezione lessicale sussume in effetti perfettamente il concetto di C-Comando e fa al caso nostro per escludere imbarazzanti casi di co-referenza attesa (in pratica i casi descritti in ii. § 4.1): un sintagma nominale (fase NP) potrà essere «atteso» (o selezionato) da una fase verbale (fase VP) all'interno della quale potrà essere lessicalizzato il clitico riflessivo «*si*», quindi «archiviato» e recuperato come possibile co-referente del riflessivo. All'interno della fase NP si potranno però attivare altre sotto-fasi (annidate), ad esempio un sintagma preposizionale o una relativa. Applicando l'idea di località della «memoria referenziale» (ogni buffer è legato alla propria fase e solo in questo buffer i referenti della fase possono essere inseriti), quello della fase VP sarà popolato solo dalle sotto-fasi immediate NP e non dalle fasi originate all'interno di questi NP. Ciò esclude la possibilità di co-referenza tra Eva e il *si* in casi come (8.c). Va notato che in questo modo non c'è bisogno di definire un «dominio locale» entro cui individuare il co-referente, come in (5), poiché questo già segue dal concetto di fase. Rispetto al caso problematico del co-referente post-verbale, (10), al momento del processamento del clitico riflessivo, il buffer referenziale non sarà popolato da nessun legittimo NP (il sintagma che lo precede è infatti un sintagma preposizionale); l'interpretazione dell'anafora deve quindi essere procrastinata fino al primo NP post-verbale utile che, una volta processato, finirà nel buffer referenziale e, successivamente, potrà completare l'interpretazione del clitico anaforico.

In conclusione, questo approccio sembra promettente per fornire una spiegazione in termini di «obliquità» del co-referente come discusso in Pollard e Sag (1994): considerando altre forme di anafora pronominale, distribuite all'interno dei diversi argomenti, la loro «obliquità» pare determinante nell'attribuzione del corretto co-referente: «Eva_i discute con Adamo_j della sua_{n/j} mela» dove Eva risulta un co-referente più marginale rispetto ad Adamo).

6. UNA RIFLESSIONE CONCLUSIVA SUL TEMPO

In base agli esperimenti con gli algoritmi di Machine Learning e con le SRN, possiamo concludere che, sulla base delle semplici evidenze distribuzionali, fenomeni quali quello della co-referenza pronominale anaforica, nel caso specifico dei riflessivi clitici in Italiano, non abbiamo individuato nessun metodo/algoritmo efficace per inferire i corretti vincoli strutturali che determinano le legittime restrizioni interpretative che qualsiasi parlante nativo dell'Italiano intuitivamente conosce. Sebbene la variabile strutturale sia quella determinante nel definire la corretta interpretazione di queste costruzioni, una necessaria asimmetria temporale, indotta da una riformulazione in chiave «antisimmetrica» della nozione di Merge e, conseguentemente, un'imple-

mentazione del concetto di «aspettativa» in termini di fasi (possibilmente annidate e/o incomplete) di processamento, pare essere il modo esplicitamente più adeguato per spiegare la maggior parte dei fenomeni qui discussi.

Concludiamo suggerendo (in modo forse speculativo) che le implicazioni computazionali di questa nozione (riconciliante) di tempo, subordinata a quella di sequenzializzazione nella costruzione di una struttura gerarchica, possano essere sfruttate in altri domini cognitivi, quali, forse, la pianificazione delle azioni: con l'obiettivo di testare, come abbiamo fatto qua, in che modo una gerarchizzazione/sequenzializzazione della pianificazione delle azioni possa essere predittiva in termini di vincoli sulle possibilità di consequenzialità tra un'azione e l'altra, si potrebbe magari tentare di caratterizzare in modo migliore le ragioni delle due posizioni inconciliabili discusse nel dibattito Pulvermüller (2014) - Moro (2014).

RIFERIMENTI BIBLIOGRAFICI

- Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., Mazzoleni, M. (2004). Introducing the La Repubblica corpus. A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. In *Proceedings of Fourth International Conference on Language Resources and Evaluation, LREC 2004*. Parigi: ELRA.
- Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43 (3), pp. 209-226.
- Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., Thornton, A.M. (2005). *Corpus e lessico di frequenza dell'italiano scritto (CoLFIS)*. Scuola Normale, Pisa.
- Bianchi, V. (2009). A note on backward anaphora. *Rivista di Grammatica Generativa*, 34, pp. 3-34.
- Chesi, C. (2012). *Competence and Computation: Toward a Processing Friendly Minimalist Grammar*. Padova: Unipress.
- Chesi, C. (2015). On directionality of phrase structure building. *Journal of psycholinguistic research*, 44 (1), pp. 65-89.
- Chesi, C., Moro, A. (2014). Computational complexity in the brain. In F.J. Newmeyer e L.D. Preston (a cura di), *Measuring grammatical complexity*. Oxford: Oxford University Press, pp. 264-280.
- Chomsky, N. (1957). *Syntactic structures*. Paris, Mouton: The Hague.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT press.
- Chomsky, N. (2008). On phases. In R. Freidin, C. Otero e M.L. Zubizarreta (a cura di), *Foundational Issues in Linguistic Theory*, Cambridge, MA: MIT Press.
- Christiansen, M.H., Chater, N. (1999). Toward a Connectionist Model of Re-

- ursion in Human Linguistic Performance. *Cognitive Science*, 23 (2), pp. 157-205.
- Christiansen, M.H., MacDonald, M.C. (2009) A usage-based approach to recursion in sentence processing. *Language Learning*, 59 (s1), pp. 126-161.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, pp. 179-211.
- Elman, J. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7 (2-3), pp. 195-225.
- Elman, J. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48, pp. 71-99.
- Fox, D., Pesetsky, D. (2005). Cyclic linearization of syntactic structure. *Theoretical linguistics*, 31 (1-2), pp. 1-45.
- Kayne, R.S. (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- McKee, C. (1992) A comparison of pronouns and anaphors in Italian and English acquisition. *Language acquisition*, 2 (1), pp. 21-54.
- Monachini, M. (1996). Elm-it: Eagles specifications for Italian morphosyntax lexicon specification and classification guidelines, *Technical Report EAG-CLWG-ELM-IT/F*, Pisa: ILC-CNR.
- Moro, A. (2014). On the similarity between syntax and actions. *Trends Cognitive Science*, 18, pp. 109-110 10.1016/j.tics.2013.11.006.
- Napoli, D.J., Sutton-Spence, R. (2010). Limitations on simultaneity in sign language. *Language*, 86 (3), pp. 647-662.
- Phillips, C. (1996). *Order and Structure*. MIT: Ph.D. Thesis.
- Pollard, C., Sag, I.A. (1994). Head-driven phrase structure grammar. Chicago: University of Chicago Press.
- Pulvermüller, F. (2014). The syntax of action. *Trends Cognitive Science*, 18, pp. 219-220, 10.1016/j.tics.2014.01.001.
- Schlenker, P. (2005). Non-redundancy: towards a semantic reinterpretation of binding theory. *Natural Language Semantics*, 13, pp. 1-92.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University press.
- Westergaard, M. (2009). Usage-based vs. rule-based learning: the acquisition of word order in wh-questions in English and Norwegian. *Journal of Child Language*, 36 (5), pp. 1023-1051.
- Witten, I.H., Frank, E., Hall, M.A., Pal, C.J. (2016). *Data Mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann.

The (apparent) divide between hierarchy and time

Abstract. Consider time as a total order among discrete events. If we look at human languages, a total order is always established among discrete events, which are the distinct pronunciations of the words (and morphemes they are formed by) in a sentence. Notice that this is not a necessary condition: if we think of the total order as a restriction imposed by the spoken modality, we could imagine a sign language in which, for instance, two arguments (e.g. «Eva» and «the apple») of a certain predicate (e.g. «eats») could be co-articulated exactly at the same time in space, one with one hand and another with the other hand in a temporally and spatially symmetrical situation. To our knowledge this pos-

sibility is limited by several restrictions in any sign language around the world. So we assume that «time» is an essential ingredient of any human language and must be included in a theory that aims at being cognitively adequate. More precisely, our hypothesis is that «time» induces an asymmetric (total) relation among spelled-out linguistic units and this is a necessary, though not sufficient, condition for the correct structural analysis of the sentence, hence for its correct interpretation. The other crucial property is hierarchy: implicit groups of words (phrase structures) that constraint temporal expectations and create complex (recursive) meaningful units. Our intent in this paper is to argue, by analyzing simple recurrent networks expectations and machine learning approaches to automatic classification on a very simple linguistic fact like anaphoric binding, that the first property (linear order) is neither a simple articulatory-perceptual restriction (also known as a «Phonetic Form effect», Fox & Pesetsky 2005) nor a sufficient distributive cue. On the other hand, we suggest that the second property (hierarchy) should be reconciled with a specific notion of time in order to solve fruitfully most of the anaphoric binding puzzles we discuss.

Keywords: anaphoric binding, syntatic structure, machine learning, simple recurrent networks.

Cristiano Chesi, Scuola Universitaria Superiore, IUSS Pavia, Palazzo del Broletto - Piazza della Vittoria n. 15, 27100 Pavia. E-mail: cristiano.chesi@iusspavia.it

Andrea Moro, Scuola Universitaria Superiore, IUSS Pavia, Palazzo del Broletto - Piazza della Vittoria n. 15, 27100 Pavia. E-mail: andrea.moro@iusspavia.it