

UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Studi Umanistici

Corso di Laurea Triennale in Lettere



MORFOLOGIA GENERATIVA E
LINGUISTICA COMPUTAZIONALE: UN
CONFRONTO TRA PARADIGMI E MODELLI

Relatore:

Prof. Francesco DEDÈ

Elaborato Finale di:

Filippo SCOTTI

Matr. n. 881504

Anno Accademico 2017/2018

Indice

Introduzione	4
Capitolo 1: la morfologia nella prima grammatica generativa	5
1.1 Morfologia in <i>Strutture della sintassi</i>	6
1.2 Morfologia in <i>Aspetti della teoria della sintassi</i>	8
1.3 Proposta trasformativale di Lees	9
1.4 Trattamento dei nominali derivati nella Teoria Standard.....	11
Capitolo 2: il primo modello di componente morfologico	13
2.1 <i>Remarks on Nominalizations</i>	13
2.2 Modello di Halle.....	15
2.3 Critiche al modello di Halle	19
Capitolo 3: la teoria morfologica di Mark Aronoff	21
3.1 Aronoff, morfologia basata sulle parole.....	21
3.2 Tre assunzioni di Scalise sulle RFP	25
3.3 <i>Morphology by Itself</i>	26
3.4 Eredità teorica del lavoro di Aronoff	27
Capitolo 4: approcci alla morfologia computazionale	29
4.1 La Linguistica Computazionale e i modelli teorici della Linguistica	30
4.2 Le applicazioni della Morfologia Computazionale	30
4.3 Trasduttore a stati finiti	32
4.4 <i>Two - Level Morphology</i>	34
Conclusioni	37
Bibliografia	39

Introduzione

Lo sviluppo della Grammatica Generativa Trasformativale, dal suo inizio fino alla metà degli anni Ottanta del secolo scorso, può essere visto nella prospettiva di una continua evoluzione del componente lessicale. Tale evoluzione è derivata non esclusivamente da considerazioni legate al lessico, ma anche da cambiamenti proposti per il componente trasformativale, il componente categoriale e il componente fonologico. Eredità di tale evoluzione fu il nuovo approccio con cui si iniziò a considerare il lessico, non più semplice lista di formativi lessicali, ma componente della grammatica dalla complessa struttura interna e luogo dove avvengono numerosi fenomeni.¹ Allo scopo di ricostruire le tappe di tale percorso si ripercorreranno sinteticamente i quadri teorici generali e le criticità in grado di stimolare mutazioni e nuove ipotesi. Il tutto avverrà nella prospettiva di un'analisi comparata tra i modelli di componente morfologico teorici (tappa finale del percorso di emancipazione della morfologia) e i modelli morfologici computazionali. Questi ultimi verranno introdotti attraverso una descrizione del loro orizzonte applicativo, imprescindibile per tutti i modelli che si definiscono 'computazionali' e quanto tali orientati al risolvimento di un problema.

Nonostante finalità differenti, i modelli teorici e computazionali per la morfologia si sono approcciati e si avvicinano alle stesse criticità di fondo. Per questo motivo, ogni passo in avanti compiuto su uno dei due versanti, quello teorico e quello computazionale/applicativo ha inevitabilmente ricadute dirette sull'altro. Si cercherà di evidenziare tale relazione attraverso l'analisi dei modelli per il componente morfologico realizzati secondo entrambe le prospettive.

¹ Cf. Scalise 1986, 17

Capitolo 1: la morfologia nella prima grammatica generativa

Nel primo capitolo si tratterà lo sviluppo dei componenti morfologico e lessicale nel quadro teorico della Grammatica Generativa Trasformativale. Tale trattazione è necessaria a ricostruire l'evoluzione storica della teoria morfologica. A questo scopo, si faranno alcuni cenni teorici sul trattamento dei problemi morfologici e lessicali in due ormai classici lavori di Chomsky, *Syntactic Structures* e *Aspects of the Theory of Syntax*² e in un volume di Robert Lees, *The Grammar of English Nominalization*. I due volumi di Chomsky, usciti rispettivamente nel 1957 e nel 1965, verranno presi in esame nella prospettiva di ripercorrere i primi approcci, nei confronti di morfologia e lessico, agli albori della Grammatica Generativa. Del volume di Lees, uscito nel 1960 e quindi precedente alla prima revisione della Grammatica Generativa Trasformativale, si prenderà in esame una proposta di formazione dei composti nominali di stampo trasformazionalista (ovvero si ipotizza che i composti nominali derivino da una struttura frasale soggiacente). Ci si concentrerà poi su alcune delle principali critiche sollevate contro di essa: come si vedrà, il trattamento trasformazionale del lessico risultò inefficiente e per questo inadeguato ad aspirare al ruolo di teoria morfologica completa e generalizzata, e sarà proprio la discussione teorica sui limiti di un modello morfologico trasformazionale a innescare i primi passi verso un'Ipotesi Lessicalista e verso l'unificazione dei meccanismi di formazione di parole all'interno del componente lessicale.

Il periodo in esame è quindi quello che va dall'anno di pubblicazione di *Strutture*, 1957, all'anno di pubblicazione di *Aspetti* 1965. Considerando l'arco temporale così delineato, oltre ai limiti del modello trasformazionale, verrà preso in esame un secondo concetto: la ridotta importanza del componente morfologico nella prima Grammatica Generativa (quella Trasformativale), a causa dalla preponderanza del componente sintattico. Nella prima GGT³ i fenomeni morfologici venivano spesso etichettati come morfo-sintattici o morfo-fonologici, associati di volta in volta al componente sintattico o

² Per i due volumi si utilizzeranno le seguenti abbreviazioni: *Syntactic Structures*: *Strutture*, *Aspects of the Theory of Syntax*: *Aspetti*.

³ Grammatica Generativa Trasformativale.

al componente morfologico. Al fine di ripercorrere per linee salienti gli sviluppi teorici del periodo, ci si è serviti in particolare della trattazione al riguardo di Sergio Scalise in *Morfologia e lessico*. Il volume, pubblicato per la prima volta in Olanda nel 1984, con il titolo *Generative Morphology*, si sviluppa proprio attorno alla questione dell'autonomia di Morfologia e Lessico nella Grammatica Generativa, questione ancora viva e attuale quando l'autore scriveva, al punto che Scalise alcuni anni più tardi nella premessa all'edizione italiana si esprime in questi termini:

In primo luogo, la morfologia non ha più bisogno di difese particolari: è diventata un campo di lavoro accreditato, serio e complesso la cui autonomia non è più in discussione. [...] In secondo luogo, oggi so che questo libro (pur con i mille difetti di tutti i prodotti umani) è un buon libro, perché – a detta di molti colleghi – ha saputo centrare i problemi dell'allora nascente campo di ricerca.⁴

1.1 Morfologia in *Strutture della sintassi*

Quando nel 1957 esce *Strutture della sintassi*, nasce formalmente la Grammatica Generativa Trasformativa, destinata a evolversi negli anni successivi (numerose le revisioni dello stesso Chomsky) sino a generare il quadro teorico della Teoria Standard. È opinione comune che, per quanto innovativa, la teoria ai suoi albori peccasse di numerosi difetti, tra i quali emergeva l'eccessivo peso assegnato alle trasformazioni, tanto che si distingue comunemente tra i primordi della teoria, definita Grammatica Generativa Trasformativa, e la successiva Teoria Standard o Grammatica Generativa, dove la componente Trasformativa ancora è presente e svolge un ruolo chiave, ma è decisamente ridimensionata nel suo utilizzo. Per questo, nonostante già in *Strutture della Sintassi* siano delineate numerose delle intuizioni di Chomsky alla base della Grammatica Generativa, è *Aspetti* il lavoro di Chomsky che 'fonda' la Teoria Standard, la quale sarà destinata a caratterizzare e orientare gran parte della ricerca linguistica nella seconda metà del novecento.⁵ Anche l'approccio nei confronti della morfologia e del lessico può essere inscritto in questa tendenza generale: dapprima ci si orientò verso proposte di tipo trasformativa, rivelatesi inadeguate e poco efficienti, e secondariamente, sulla scorta

⁴ Scalise 1990, 9.

⁵ Cf. Robins 1997, 132.

delle critiche al trattamento trasformatzionale di lessico e morfologia, si adottarono nuovi modelli nel quadro della Teoria Standard.

Come è noto Chomsky muove i primi passi criticando la linguistica strutturale, da lui accusata di essere eccessivamente descrittiva ed incapace di dar conto della creatività del linguaggio. Ma dagli strutturalisti eredita il rigore formalista e le rappresentazioni ‘matematiche’ dei fenomeni linguistici. Anche in *Strutture della Sintassi*, la struttura dei costituenti di una frase viene rappresentata tramite diagrammi ad albero, ovvero gli indicatori sintagmatici. Le regole che si occupano dei passaggi da un livello all’altro nel diagramma, appartengono al componente categoriale e si chiamano regole di riscrittura. Troviamo due tipi di regole di riscrittura, le regole di struttura sintagmatica e le regole di inserzione lessicale, le quali agiscono in ultima istanza rivelando la struttura superficiale della frase, passando dai costituenti alle voci lessicali. È evidente come manchino regole specifiche del componente lessicale, differenziate da quelle del componente categoriale.⁶ Ne risulta che questa proposta non tiene conto di problemi strettamente lessicali, legati a regole specifiche delle voci inserite. Ad esempio: un verbo intransitivo non può essere seguito da un oggetto (1a), oppure, un verbo non può avere come soggetto un oggetto inanimato(1b).

- (1)(a) Marco ride il bicchiere
- (b) Il bicchiere beve Marco

La non grammaticalità delle due frasi deriva da caratteristiche proprie delle voci lessicali, in (1a) la intransività del verbo cui segue un complemento oggetto e in (1b) il fatto che ‘bicchiere’ sia un oggetto inanimato. Scalise evidenzia come in *Strutture* frasi come quelle nell’esempio non costituissero un problema. Non era scopo della nascente teoria realizzare una grammatica completa. Inoltre, Chomsky suggeriva come possibile soluzione di imporre diverse restrizioni sulla scelta del Verbo in relazione a soggetto e oggetto in modo da evitare (1b), e per risolvere l’errore di (1a), di diversificare le regole di riscrittura per il SV: una per i verbi transitivi ed una per i verbi intransitivi. Successivamente, in *Aspetti*, la soluzione ‘sintattica’ verrà abbandonata dallo stesso Chomsky a favore di una soluzione ‘lessicale’.

⁶ Cf. Scalise 1986, 17.

1.2 Morfologia in *Aspetti della teoria della sintassi*

In *Aspetti* il lessico e le regole di riscrittura vengono separati. Chomsky rileva che molte proprietà dei formativi, irrilevanti per il funzionamento delle regole di base e spesso idiosincratiche, possono essere specificate nel lessico, alleggerendo notevolmente la grammatica e procede secondo questa linea: “In general, all properties of a formative that are essentially idiosyncratic will be specified in the lexicon⁷”. Scalise propone come esempio la gestione dei verbi transitivi: è possibile aggiungere alle voci lessicali verbali un tratto, il quale permetta di distinguere quali sono i verbi transitivi che ammettono la cancellazione dell’oggetto (*leggere, mangiare*) e quali sono i verbi transitivi che non ammettono la cancellazione dell’oggetto (*spaventare, mettere*). Così solo alla prima classe si applica la regola trasformazionale che cancella l’oggetto e non vengono generate frasi scorrette.

L’assegnazione di proprietà alle voci lessicali è un passo verso una maggiore profondità ed autonomia del componente lessicale. Nella Teoria Standard ogni entrata lessicale viene accompagnata da informazioni sintattiche, semantiche e fonologiche e vengono delineate tutte le proprietà idiosincratiche non prevedibili con regole generali. Scalise elenca alcune delle proprietà sintattiche⁸ che in *Aspetti* vengono associate alle entrate lessicali:

- (2)(a) categoria lessicale
- (b) tratti inerenti
- (c) tratti contestuali
 - (c - i) sotto-categorizzazione stretta
 - (c - ii) restrizioni selettive

La prima informazione indica naturalmente la categoria grammaticale, per cui il tratto [+N] significherà nome, il tratto [+V] significherà verbo eccetera. I tratti inerenti specificano proprietà della parola di tipo semantico (vedi nota 2), come [+animato],

⁷ Chomsky 1965, 126.

⁸ Lo stesso Scalise specifica che se la natura di questi tratti sia ‘sintattica’ o ‘semantica’ è materia controversa.

[+umano] oppure [+astratto]. Si noti che specificare questi tratti risolve problemi come quello nell'esempio sopra riportato (1a): un nome con tratto [-animato] non potrà essere associato all'azione descritta dal verbo 'bere'. I tratti contestuali di sotto-categorizzazione stretta ci permettono di specificare ad esempio se un nome può essere preceduto o meno dall'articolo. Infine, le restrizioni selettive, specificano quali tratti devono possono avere altri elementi nel contesto. Ovvero, un tratto di sotto-categorizzazione stretta specifica se un verbo può trovarsi davanti a SN (sintagma nominale), un tratto di restrizione selettiva indica quali tratti deve possedere quel SN.

Possiamo riassumere che secondo Chomsky la voce lessicale deve specificare: aspetti della struttura fonetica non prevedibili con una regola generale, proprietà pertinenti al funzionamento delle regole trasformative, proprietà del formativo pertinenti all'interpretazione semantica e tratti lessicali indicanti le posizioni in cui un formativo lessicale può essere inserito. In aggiunta le proprietà vengono specificate in termini di regole di ridondanza, per cui regole generali stabiliscono quando un tratto categoriale ne implica un altro (per esempio [+umano] implica [+animato] è stabilito da una regola generale).

È da rilevare anche l'aggiunta della regola di Inserzione Lessicale, assente in *Strutture*. Attraverso questa regola le voci lessicali vengono inserite al posto dei simboli inoperanti nella struttura profonda, tenendo conto sia degli stessi simboli, sia della natura del contesto. Nel suo complesso la proposta avanzata nella Teoria Standard arricchisce il componente del lessico, attribuendo alle voci lessicali molte più informazioni grammaticali Cf. Scalise, 24.

1.3 Proposta trasformativa di Lees

Nel sistema fin qui delineato le voci ammesse nel lessico erano solamente parole "semplici" escludendo parole composte e parole derivate. Nel 1960 Lees avanzò una proposta trasformazionalista per la costruzione dei composti nominali. Lees ipotizza che i composti nominali possano essere generati da una serie di trasformazioni a partire da una struttura frasale profonda. All'interno della struttura frasale originaria le relazioni grammaticali tra i formativi del composto nominale vengono espresse esplicitamente. Lees propone tre argomenti a sostegno dell'ipotesi:

- La comprensione dei composti nominali avviene attraverso relazioni sintattiche tra i formativi, queste relazioni possono essere descritte come relazioni tra costituenti
 - L'ipotesi trasformativale giustifica l'ambiguità del significato di alcuni composti: questa può dipendere da diverse strutture soggiacenti alle quali corrispondono diversi significati del composto
 - Esistono casi in cui la differenza tra relazioni grammaticali è annullata in superficie, le strutture profonde tengono conto di questa differenza.
- (3) (a) *Windmill* : mulino a vento → mulino azionato dal vento
 (b) *Flour mill* : mulino per la farina → mulino che macina la farina

Una delle critiche mosse alla proposta di Lees riguarda la cancellazione del materiale lessicale. In *Windmill* viene cancellato il verbo 'azionare' ed è possibile fare numerosi esempi di composti nei quali un verbo esistente in una ipotetica struttura soggiacente viene cancellato. Il problema di queste cancellazioni è che richiedono trasformazioni aggiuntive, in contrasto con la nozione di "linguaggio naturale". Per questo in *Aspetti* viene proposto un principio di recuperabilità delle trasformazioni, il quale restringe le cancellazioni a elementi in qualche modo presenti nella struttura frasale (come il tu nell'imperativo).

Un secondo problema è l'interpretazione data dei composti, i quali possono avere più di un significato. La struttura frasale di (3a) è 'il vento aziona il mulino' ma dal momento che il verbo è stato cancellato si potrebbe proporre altre parafrasi e ogni parlante potrebbe interpretarlo in maniera diversa. Esistendo un "rango di possibili parafrasi" bisognerebbe associare regole di trasformazione ad ognuno di essi.

Un terzo problema sono i composti il cui significato non si basa unicamente sulla struttura profonda. Scalise propone come esempio la frase '*A green black-board*', facilmente interpretabile e accettabile per un parlante, mentre la struttura soggiacente dovrebbe essere '*a green board wich is black*', sicuramente non accettabile. Vi sono poi numerose idiosincrasie delle quali regole trasformativazionali non posso dar conto, poiché, come sostiene Scalise (1990, 33): "non v'è modo in cui un composto derivato trasformativionalmente possa acquisire tratti idiosincratici". Più in generale, il lessico è un campo prettamente idiosincratico, ed è molto difficile trattarlo con regole generali.

Tramite trasformazioni è possibile generare numerosi composti e derivati possibili, ma non esistenti nella lingua. Il concetto di frase possibile, come metro di correttezza, esiste solo nel campo della sintassi. Una parola derivata può essere possibile ma allo stesso tempo essere sbagliata, ovvero non esistere. Un trattamento trasformativale richiederebbe regole che blocchino tutte le trasformazioni che generano parole non esistenti. Trattare le idiosincrasie all'interno del lessico è maggiormente performante. Per esempio, all'interno di un indicatore sintagmatico non si dovrebbe dar conto della struttura soggiacente di un composto e sarebbe sufficiente etichettarlo come costituente di categoria N.⁹

1.4 Trattamento dei nominali derivati nella Teoria Standard

Come si è visto, diverse furono le obiezioni sollevate contro la proposta di Lees. Allo stesso modo, quando all'interno della Teoria Standard fu proposto un trattamento trasformativale della formazione dei derivati nominali, vennero sollevate diverse critiche. Similarmente al trattamento dei composti proposto da Lees, il trattamento trasformativale dei derivati associa ai sintagmi derivati una struttura soggiacente. Scalise fa il seguente esempio:

(1) La critica del libro → X critica il libro

Nel caso mostrato si assume che la voce verbale sia la forma base, ed il nome la forma derivata. È da notare che questo trattamento si iscrive agilmente nel quadro trasformativale, ove uno dei principali paradigmi metodologici consiste nell'individuare forme base da cui derivare nuove forme tramite regole. Un vantaggio di questo trattamento è la possibilità di associare ai nominali corrispondenti le specifiche selezioni restrittive dei verbi. Come si è visto in precedenza, in *Strutture*, frasi non grammaticali dove un soggetto inanimato è associato a un verbo che può avere solo soggetti animati, sono scartate imponendo restrizioni selettive sul soggetto del verbo. Ad esempio, il verbo *Rifiutare* descrive un'azione che non può essere compiuta da un soggetto inanimato, come può essere *Tavolo*. Estendendo le restrizioni selettive sul soggetto del verbo *Rifiutare*, al nominale derivato *Rifiuto*, si scongiurano situazioni di non grammaticalità come *Il rifiuto*

⁹ Nome.

del tavolo.¹⁰ Andrew Spencer, in *Morphological Theory*, descrive così l'orientamento secondo cui evolve inizialmente la Teoria Standard:

One path was to stress the importance of transformations and use this formal device to express as many relationships between linguistic forms as possible. This led to the appearance of Generative Semantics.¹¹

Secondo Spencer, la caratteristica più significativa di questo orientamento è che viene abolita la struttura profonda, o meglio, viene identificata con le strutture semantiche. È il caso del trattamento trasformatzionale dei composti e dei derivati visti in precedenza. Spencer constatò inoltre, che, per quanto concerne la morfologia flessionale, i morfemi venivano considerati come diverse realizzazioni fonetiche dovute alla struttura sintattica. Nel complesso, i fenomeni morfologici erano sempre trattati in associazione al componente sintattico o al componente fonologico:

The major phenomena of morphology seemed to come under the purview of syntactic and phonological transformations and the effects of generative semantics seemed to reinforce the trend towards making the major phenomena of morphology solely the responsibility of syntax and phonology.¹²

Il primo passo decisivo, nella direzione di una morfologia autonoma e di rilievo nella Grammatica Generativa, fu motivato proprio dalla necessità di superare i limiti di un modello esclusivamente trasformatzionale. Il primo a intraprendere con decisione questa direzione sarà Chomsky, con il suo lavoro *Remarks on Nominalizations* (1970), nel quale critica il trattamento trasformatzionale dei derivati e fonda la cosiddetta Ipotesi Lessicalista.

¹⁰ Cf. Scalise 1984, 38.

¹¹ Spencer 1991, 67.

¹² *ibid.*

Capitolo 2: il primo modello di componente morfologico

In questo capitolo si vedranno in particolare il lavoro di Chomsky *Remarks on Nominalizations* e il cosiddetto modello di Halle, esposto nell'articolo *Prolegomena to a Theory of Word Formation*. I limiti del modello trasformatzionale non saranno più al centro dell'attenzione e si discuteranno i primi approcci al componente morfologico in un quadro lessicalista, ovvero dopo l'affermazione della necessità di un componente autonomo per lessico e morfologia. Anche se il quadro teorico, le soluzioni e i paradigmi sono notevolmente cambiati rispetto a quelli qui presentati (pubblicati negli anni '70), i problemi e le questioni al centro della ricerca morfologica sono rimasti in gran parte gli stessi.

2.1 *Remarks on Nominalizations*

In *Remarks on Nominalization* Chomsky critica il trattamento trasformatzionale dei derivati elaborato all'interno della Teoria Standard (v. *supra*, § 1.4). Nella Teoria Standard esisteva una naturale tensione a trattare ogni tipo di relazione tra due parole come una trasformazione da una forma base ad una forma derivata (come da paradigma nel trasformazionalismo). Conseguenza principale della critica impostata da Chomsky, fu una riaffermazione teoretica del carattere idiosincratico del lessico. Nel suo lavoro, Chomsky, motiva le sue obiezioni sulla base delle differenze esistenti tra “nominali gerundivi” e “nominali derivati”. Vengono individuate tre principali differenze: la produttività sintattica, la regolarità semantica e la struttura interna dei nominali. Per quel che concerne la produttività sintattica, Chomsky osserva che con quasi tutti i nominali gerundivi si possono formare frasi, ma lo stesso non si può dire per i nominali derivati. Si propone l'esempio di Scalise (1984, 38):

- (2) It is easy to please John : forma base
 - Its being easy to please John : gerundivo
 - The easiness to please John : derivato
- It seems that John is gone: forma base
 - Its seeming that John is gone: gerundivo

Per quanto concerne la regolarità semantica, Chomsky rileva che mentre il significato dei nominali gerundivi è regolare, in quanto prevedibile a partire dal verbo, il significato dei nominali derivati è spesso idiosincratico. Alcuni esempi proposti da Chomsky sono nomi come *Marriage*, *Revolution*, *Construction*, dove nel corso della derivazione dal verbo corrispondente vi è un mutamento sostanziale del materiale semantico. La terza differenza è quella attinente alla struttura interna: in un indicatore sintagmatico un nominale derivato viene indicato come SN¹³, mentre un nominale gerundivo viene indicato come F.¹⁴ Conseguenza logica è che solo SN può avere il plurale: *John's many refusals of the offer* è grammaticale, *John's many refusings of the offer*, non lo è. Un'ulteriore differenza causata dalla struttura interna è che i nominali derivati vengono modificati da un aggettivo, i nominali gerundivi da un avverbio.¹⁵

Sulla base delle differenze sopra delineate, Chomsky esclude che i nominali derivati possano essere generati con regole trasformazionali a partire da verbi sottostanti. Nelle parole di Spencer:

The primary importance of this paper¹⁶ for morphology was that it pointed to the need for a separate theory of derivational morphology, distinct from the theory of syntactic transformations.¹⁷

Remarks on Nominalizations segnò il passaggio verso un'Ipotesi Lessicalista nel quadro teorico della Grammatica Generativa. Chomsky propose infatti di trattare all'interno del componente lessicale i nominali derivati, tramite, ad esempio, regole morfologiche. In *Remarks* vengono considerati solamente i nominali derivati, ma le conseguenze teoriche si estesero all'intero trattamento del lessico: l'utilizzo delle trasformazioni fu consentito solo per il trattamento di processi regolari e privi di idiosincrasie e dall'Ipotesi Lessicalista di Chomsky, limitata ai nominali derivati, si passò alla Ipotesi Lessicalista Estesa, così formulata da Jackendoff:

¹³ Sintagma nominale.

¹⁴ Frase.

¹⁵ Cf. Scalise 1984, 39.

¹⁶ Riferito a *Remarks on Nominalizations*.

¹⁷ Spencer 1991, 67.

Transformations cannot change node labels, and they cannot delete under identity or positive absolute exception [...]. This means that only changes that transformations can make to lexical items is to add inflectional affixes such as number, gender, case, person and tense. Transformations will thus be restricted to movement rules and insertion and deletion of constants and closed sets of items.¹⁸

Scalise rileva come l'Ipotesi Lessicalista estesa sia stata un passo fondamentale verso un'ipotesi 'modulare' della organizzazione della grammatica, dove i 'moduli' interagiscono nella formazione delle frasi ognuno con proprie regole specifiche. In questo modo il componente trasformativale non dovette più trattare problemi per i quali risultava inadeguato. È un passaggio d'importanza cruciale verso la nascita di un componente morfologico autonomo.

2.2 Modello di Halle

Prolegomena to a Theory of Word Formation è il titolo di un articolo pubblicato da Morris Halle nel 1973. Halle era stato a lungo collaboratore di Chomsky, in particolare nello sviluppo della fonologia generativa. Nell'articolo, lavorando nel solco dell'Ipotesi Lessicalista, Halle propone un modello di componente morfologico autonomo. Nelle parole di Scalise, *Prologomena*, articolo molto breve e di carattere programmatico, contribuì "alla fondazione e alla ispirazione di tutta la moderna ricerca morfologica nel quadro teorico generativo" (Scalise 1984, 46). Halle partì dall'osservazione che se una grammatica rappresenta formalmente la conoscenza linguistica di un parlante della propria lingua, allora deve esistere una componente di essa che giustifichi la conoscenza del lessico. Lo scopo del modello di Halle è di rappresentare la conoscenza lessicale del parlante descrivendo questa componente.

L'impostazione del lavoro è programmatica, vengono individuate tre domande cui una teoria grammaticale dovrebbe dare una risposta. Ovvero, come può la grammatica codificare:

- (1) l'inventario delle parole esistenti in una lingua?
- (2) l'ordine in cui i morfemi appaiono nelle parole?

¹⁸ Jackendoff 1972, 13.

(3) le caratteristiche idiosincratiche delle parole individuali?

La prima domanda dà per assunto che ci riferiamo al parlante ideale di una lingua. Questo significa che non si cerca di dare conto di quale sia il vocabolario completo di una lingua, né di quale porzione di esso i parlanti siano a conoscenza. A livello teoretico, la domanda si riferisce a come avviene la distinzione tra una parola esistente ed una parola potenziale di una lingua.¹⁹ La seconda domanda nasce dall'osservazione seguente: i parlanti di una lingua conoscono l'ordine in cui i morfemi possono apparire in una parola. Un parlante ideale inglese sa che la parola *transformational* si divide nei morfemi *trans* + *form* + *at* + *ion* + *al* e sa che qualsiasi altro ordine di questi morfemi sarebbe scorretto. Spencer puntualizza che esistono lingue, specialmente quelle a forte carattere agglutinante, in cui i morfemi possono apparire in diverso ordine, dando luogo a mutazioni nel significato. Questo comporta che la grammatica non deve solamente dare conto di quale sia il corretto ordine dei morfemi, ma anche di quali variazioni comportino nel significato ordini differenti. La terza domanda di Halle centra una delle questioni già trattate negli scorsi paragrafi, ovvero il carattere idiosincratico del lessico. Halle introduce la questione prendendo spunto proprio dalla trattazione dei nominali derivati di Chomsky. A tutte le domande Halle fornisce una risposta elaborando il seguente sistema:

¹⁹ Cf. Spencer 1991, 74.

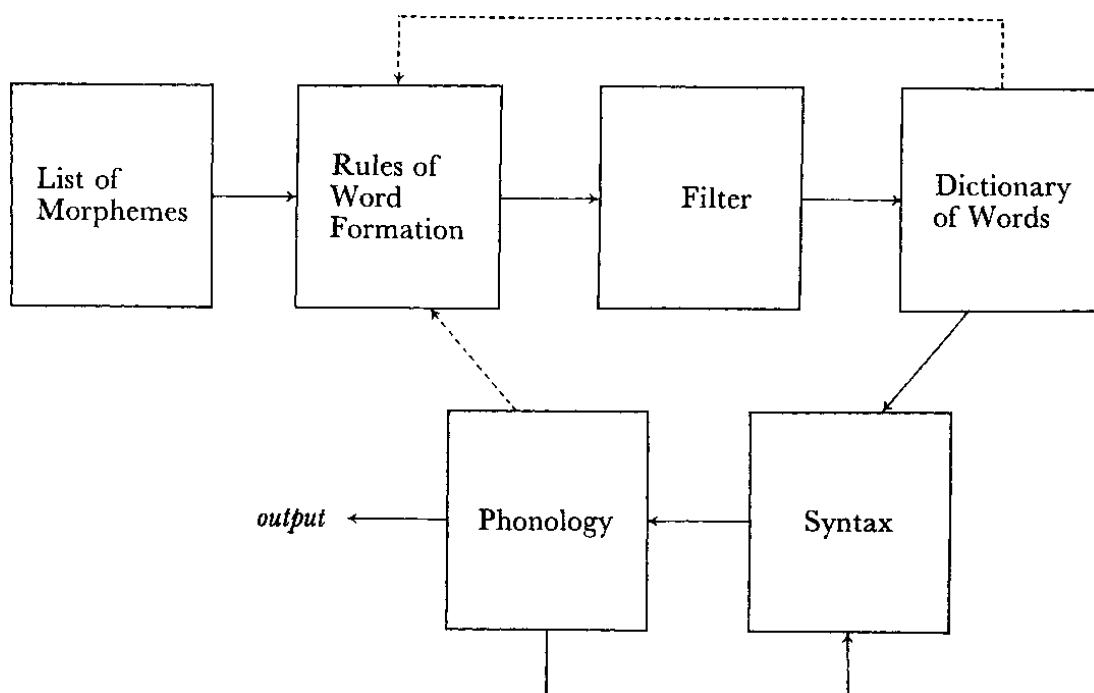


Figura 1 Modello del componente morfologico. Halle (1973, 8).

Alla base del modello vi è l'assunzione che il lessico consista in una lista di morfemi ordinati da un set di regole, definite Regole di Formazione di Parola (RFP). I morfemi vengono registrati come segmenti fonologici ed etichettati secondo la categoria sintattica ed alcune proprietà grammaticali (ad esempio: *write* viene etichettato come verbo e in aggiunta possiede il tratto [-latino] poiché appartiene alla parte 'non latina' dei morfemi ed il tratto [+forte] poiché appartiene al gruppo dei verbi forti inglesi²⁰). Le RFP ordinano i morfemi secondo regole generali, e non generano solamente le parole 'esistenti', ma anche quelle 'possibili' di una data lingua. Esistono due tipologie di RFP, quelle che si applicano alle basi (stems) e quelle che si applicano alle parole.

Ogni RFP registra le categorie sintattiche delle uscite e solo alcune forniscono anche informazioni semantiche, se queste sono adatte ad essere trattate con regole generali. Ad esempio, in inglese il suffisso *-hood*, in parole come *boyhood*, segna un cambio di tratto da [-astratto] a [+astratto].

Siccome le regole di formazione di parola non generano solo le parole esistenti di una lingua, ma anche tutte le parole 'potenziali', solo le parole esistenti vengono codificate all'interno del Dizionario, il quale è collegato al componente delle RFP in uscita e in

²⁰ Cf. Scalise 1984, 48.

entrata, in quanto riceve le parole generate secondo le RFP e fornisce a sua volta parole sulle quali applicare RFP. Un ruolo chiave viene svolto dal componente chiamato Filtro, intermediario tra le RFP in uscita e il Dizionario: è il filtro a gestire la componente idiosincratica del lessico, applicando le dovute correzioni e aggiungendo i tratti non prevedibili (e quindi non specificabili tramite regole). È possibile quindi distinguere tra formazioni di parole regolari e formazioni di parole idiosincratiche, l'insieme delle parole formate secondo questi due meccanismi confluisce nel Dizionario.

Halle cerca di rendere conto anche di come la fonologia influisca sulle regole morfologiche. Due sono i modi possibili secondo cui la fonologia influenza la formazione di parole: un morfema può subire delle restrizioni a combinarsi solo con morfemi con una specifica forma morfologica. È questo il caso di affissi allomorfi che vengono selezionati in dipendenza dalla parte finale della base. Ad esempio, in ungherese la seconda persona singolare dei verbi al presente può essere segnalata da due diversi affissi: *-ol* dopo una sibilante ed *-(a)sz* negli altri casi. Condizionamenti fonologici come questo possono essere gestiti da regole generali in quanto prevedibili. Un secondo tipo di condizionamento fonologico avviene quando le regole fonologiche modificano una parola prima delle RFP, variando le condizioni in cui la RFP può essere applicata. Ad esempio, in inglese, l'affisso incoativo/causativo *-en*, il quale si attacca agli aggettivi e genera verbi che significano 'diventare, causare di essere quell'aggettivo', apparentemente subisce un condizionamento fonologico del primo tipo, e può essere associato solo a basimono sillabiche che terminano con una ostruente, a volta preceduta da una sonora. Come nei casi:

(4) Quicken

Redden

*Slowen → mancante

*Greenen → mancante

Esistono però casi in cui questa restrizione sembra essere violata, come in *soften*, *fasten*, dove *-en* si attacca a due ostruenti. La restrizione non viene realmente violata, poiché interviene una regola fonologica che elimina la /t/ in queste parole e questo significa che la RFP viene applicata solo successivamente alla regola fonologica. La

precedenza di una regola fonologica su di una RFP rende necessario un ripensamento dell'ordine di intervento delle regole nella grammatica. Nel modello di Halle le RFP precedono la sintassi, e nella Teoria Standard la sintassi precede le regole fonologiche.²¹ Ne consegue che, per rendere funzionante il modello di Halle, la fonologia deve rinviare le parole alle RFP, come illustrato sopra nella *figura 1*. La possibilità di generare *loop* e il continuo scambio di forme tra i diversi componenti è la principale differenza tra le RFP e le vecchie regole trasformazionali, costruite sul paradigma sequenziale ad impostazione 'forma base - forma derivata'.

Le Regole di Formazione di parola furono adottate e rielaborate negli anni successivi, orientando in questo modo la successiva ricerca morfologica.²² Pochi anni più tardi Aronoff propose di circoscrivere le RFP alle sole regole di tipo 'parola + morfema', fondando la teoria della 'morfologia basata sulle parole', che illustreremo brevemente nel prossimo capitolo.

Da un punto di vista teoretico il lavoro di Halle fu rilevante poiché ipotizzava un componente autonomo e unitario, il lessico, dove trattare i fenomeni morfologici e un insieme di regole le RFP, incaricate di rendere conto della formazione delle parole. Lavorando sulla formazione delle parole, Halle, evidenziò una differenza fondamentale tra sintassi e morfologia, i concetti di 'possibile' ed 'esistente'. Dove nella sintassi i due concetti coincidono, indicando frasi possibili e rendendo conto della produttività del linguaggio, nel lessico solo le parole 'esistenti' possono essere giudicate corrette.²³

2.3 Critiche al modello di Halle

Il modello di Halle collocò la morfologia al centro del dibattito scientifico e ispirò numerose successive ricerche, ma non fu esente da critiche e rielaborazioni, in particolare per quanto concerne quattro difetti:

- (a) il modello è orientato sui morfemi e non sulle parole
- (b) eccesso di potere delle RFP
- (c) non finitezza del Filtro

²¹ Cf. Spencer 1991, 77-78.

²² Cf. Scalise 1986, 51.

²³ Cf. *ivi*, 57.

- (d) ridondanza del sistema derivante dai due livelli di rappresentazione postulati (la lista dei morfemi e il dizionario).

Halle progettò il suo modello basandosi principalmente sulla lingua inglese, dove parole e morfemi coincidono molto spesso. Lo stesso non si può dire per lingue con strutture morfologiche più complesse. Inoltre, nella lista dei morfemi non vi era distinzione tra affissi derivazionali/flessivi e parole, oscurandone le differenze. I vantaggi di una morfologia basata sulle parole verranno indicati meglio nel prossimo capitolo, illustrando la teoria di Aronoff.

Un problema delle RFP è che non sono sufficientemente ristrette e possono operare anche sull'insieme delle parole presenti nel dizionario. Regole di questo tipo aumentano enormemente il potere della grammatica, ed è difficile stabilire, per esempio, dopo quante derivazioni una parola non possa più essere soggetta a RFP. Allo stesso tempo molte delle parole formate da RFP non possono essere escluse né sintatticamente né fonologicamente, ad esempio: il suffisso *-al* può essere aggiunto a nomi astratti e comparire in sequenze come *baptismal*, ma non può essere utilizzato per formare parole come *dogmatismal* o *fatalismal*, le quali rientrano nell'insieme delle parole potenziali ma non esistenti della lingua inglese.²⁴

Si obietta al componente del Filtro di essere incompleto, siccome l'insieme delle parole possibili ma non esistenti non è a sua volta un insieme finito. È stato constatato²⁵ che non è possibile porre un limite grammaticale al grado di complessità di parole derivate e composte e che a determinare la massima lunghezza di un composto sono questioni esecutive, come la memoria di un parlante.

Si è già rilevato come dizionario e lista dei morfemi si sovrappongono parzialmente, introducendo un grado di ridondanza che riduce l'efficienza del sistema.

Diverse furono le questioni ancora aperte dopo la pubblicazione del modello di Halle, il suo merito è quello di aver centrato le questioni chiave del modello allora nascente, spostando definitivamente la discussione morfologica dal campo trasformatzionale a quello lessicalista.

²⁴ Cf. *ivi*, 59.

²⁵ Booij, 1977.

Capitolo 3: la teoria morfologica di Mark Aronoff

In questo capitolo si passerà in rassegna il lavoro di Mark Aronoff, linguista le cui ricerche sono incentrate sulla morfologia e sulle sue relazioni con gli altri componenti della grammatica. Il metodo utilizzato sarà lo stesso dei due capitoli precedenti, si ripercorreranno proposte e modelli teorici per linee salienti, avvalendosi sia di sintesi teoriche redatte da terzi (i volumi di Scalise e Spencer già citati in precedenza) sia del confronto diretto con il lavoro di Aronoff e in particolare dei tre volumi *Word Formation in Generative Grammar* (1976), *Morphology by Itself* (1994) e *What is Morphology* (2005) (di quest'ultimo Aronoff è coautore assieme a Kirsten Fudeman). Se a Chomsky è attribuibile il merito di avere per primo intuito la necessità di un componente morfologico autonomo nel contesto generativo e ad Halle quello di avere per primo sistematizzato questa intuizione traducendola nella proposta di un modello completo, ad Aronoff si può riconoscere il merito di avere definitivamente consolidato, anche attraverso sensibili innovazioni rispetto ad Halle, il campo di ricerca della morfologia generativa.

3.1 Aronoff, morfologia basata sulle parole

Si è già accennato in precedenza al fatto che una delle innovazioni di maggior rilievo introdotte nelle teorie di Aronoff fu il concetto di morfologia basata sulle parole. Questa teoria viene sviluppata all'interno di *Word Formation Grammar* (1970). Aronoff espone così il principio chiave della teoria:

All regular word-formation processes are word-based. A new word is formed by applying a regular rule to a single already existing word. Both the new word and the existing one are members of major lexical categories.²⁶

La prima conseguenza di questa assunzione è che le RFP debbano lavorare esclusivamente sulle parole e non sui morfemi. Le implicazioni sono sia di tipo pratico sia di tipo teoretico e possono essere riassunte come segue (i punti sono stati elaborati

²⁶ Aronoff 1976, 21.

originariamente da Scalise, qui vengono ripresi da Spencer (1991, 83), il quale adottò la lista di Scalise con alcune modifiche:

- (1) La base delle RFP sono le parole.
- (2) Queste parole devono essere parole esistenti. Una parola possibile ma non esistente non può essere la base di una RFP.
- (3) Le RFP possono avere come base una singola parola, non di più (frasi).
- (4) L'input e l'output di una RFP deve essere un membro delle categorie N, V, A.²⁷

Anche se non più alla base delle RFP e dei processi generativi, i morfemi svolgono ancora un ruolo di rilievo, ovvero quello di dar conto della struttura interna di una parola (come si vedrà più avanti). La prima giustificazione utilizzata da Aronoff a sostegno della sua tesi è che la morfologia derivazionale sembra non operare su altro che le parole. È da notare come Aronoff non consideri nel suo sistema la morfologia flessionale, in quanto troppo condizionata dalla sintassi (adotta quindi la posizione della Teoria Standard dove la morfologia flessionale rientra nel campo misto della morfo-sintassi²⁸). Per dimostrare che solo le parole sono alla base dei processi derivazionali si imposta una riflessione sul concetto di segno minimo. Chomsky (1970) stabilì che il segno minimo della sintassi fosse la parola, mentre in termini più generali (anche all'esterno del quadro generativista) il morfema veniva definito come l'unità più piccola del linguaggio ad avere significato. Aronoff osserva che un segno minimo debba essere stabilito in relazione al livello di analisi del linguaggio in uso e sostiene che, nel caso del componente morfologico, il segno minimo debba essere la parola. A squalificare il morfema come segno minimo della morfologia è l'esistenza di morfemi privi di un significato regolare e indipendente al di fuori delle parole in cui ricorrono.²⁹

This is the basic trouble with morphemes. Because words, though they may be formed by regular rules, persist and change once they are in the lexicon, the morphemes out of which words seem to have been formed, and into which they seem

²⁷ Nome. Verbo, Aggettivo.

²⁸ Cf. Spencer 1991, 85.

²⁹ Cf Scalise 1984, 63-64.

to be analyzable, do not have constant meanings and in some cases, they have no meaning at all. It is this persistence which forces us to adopt a lexicalist hypothesis.³⁰

Anche nel modello di Halle viene trattato il problema della incertezza semantica dei morfemi. La soluzione proposta è il componente del Filtro, chiamato a dar conto di tutte le proprietà fonologiche, sintattiche, semantiche non derivabili a partire da regole generali. Ma secondo Aronoff, un componente come il filtro non è adeguato a rendere conto di mutamenti semantici radicali, come nel caso della parola *trasmissione*, che secondo le regole dovrebbe essere un nominale astratto, ma può avere anche un significato diverso quando riferito alla parte meccanica di un'automobile.³¹ Questo genere di mutamenti è molto frequente nel linguaggio e nel corso del tempo una parola esistente può acquistare nuovi usi. A volte si tratta di fenomeni di polisemia, ma in alcuni casi la diversità dei significati è tale da dover considerare l'esistenza di due parole omonime. Ne consegue che la morfologia è adatta a spiegare il significato e la formazione delle sole parole regolari, mentre il lessico deve occuparsi delle idiosincrasie elencando tutte le parole esistenti, regolari e non.

Un secondo problema è quello dei morfemi che non possiedono alcun significato. Aronoff prende in esame per la sua dimostrazione i nomi di alcune bacche in inglese:

- (1) Cranberry
- Boysenberry
- Huckleberry

La parola *berry* in inglese significa bacca e le tre parole possono essere analizzate come l'insieme di *berry* e rispettivamente i tre morfemi:

- (2) Cran
- Boysen
- Huckle

³⁰ Aronoff 1976, 18.

³¹ Cf. Scalise 1984, 69.

Nella lingua inglese nessuno di questi tre morfemi ricorre come parola o come morfema all'interno di combinazioni differenti da quelle elencate in (1), quindi si può affermare che nessuno dei tre possiede un significato indipendente al di fuori della parola in cui compare. Aronoff puntualizza che questo argomento può essere esteso a molte parole del lessico inglese. Come esempio propone un gruppo di parole inglesi formate da radici latine e affissi:

X-fer	X-mit	X-sume	X-ceive	X-duce
refer	Remit	resume	Receive	reduce
defer	Demit		Deceive	Deduce
confer	Commit	consume	Conceive	Conduce
re-X	con-X	in-X	de-X	
repel	Compel	impel		
remit	Commit		Demit	
refer	Confer	Infer	Defer	

In una teoria in cui i morfemi sono il segno minimo con significato dovremmo trovare il valore semantico di tutti gli affissi e di tutte le radici. Ma nei casi sopra elencati il valore semantico non è costante. Il significato di *fer*, *mit*, *sume*, *ceive*, *duce*, non è lo stesso all'interno di parole diverse e in alcuni casi cambia completamente da parola a parola, tanto che non è possibile trovare alcuna similarità sul piano semantico.³²

³² Si potrebbe supporre che non si tratti dello stesso morfema bensì di allomorfi. Aronoff scarta questa ipotesi osservando che gli *stems* hanno tratti in comune e appartengono tutti alla classe dei verbi con radice latina.

3.2 Tre assunzioni di Scalise sulle RFP

Scalise illustra con tre assunzioni il funzionamento delle RFP nella teoria di Aronoff. La prima assunzione è che la principale differenza tra le RFP di Aronoff e le RFP di Halle, sia che le RFP di Aronoff non hanno accesso agli altri componenti della grammatica e funzionano esclusivamente all'interno del componente lessicale (mentre si è visto che in Halle le RFP dialogavano con il componente fonologico). Questo non significa che le RFP non possano essere applicate in riferimento a proprietà sintattiche, semantiche, morfologiche o fonologiche, ma che alla loro base vi sono solamente le parole presenti nel dizionario e non modificate da regole diverse dalle RFP.

La seconda assunzione di Scalise è che le RFP differiscono dalle regole degli altri componenti nel loro carattere facoltativo. Le altre regole (sintattiche e fonologiche) sono necessarie per la creazione di una frase, viceversa è possibile generare una frase senza utilizzare alcuna RFP (sarebbe sufficiente creare la frase con sole parole semplici prese dal dizionario). Per questo motivo Aronoff esclude che la RFP associata alla generazione di una parola complessa venga applicata dal parlante ogni volta che utilizza tale parola e definisce le RFP *once only rules*, ovvero suppone che il parlante utilizzi la RFP per generare la parola complessa solo la prima volta, per poi immagazzinare la parola nel dizionario.

La terza assunzione di Scalise è riferita alla differenza tra i concetti di “parola” e di “affisso”. Si è visto che nel modello di Halle i morfemi (e quindi anche gli affissi) vengono elencati nella lista dei morfemi. La lista è assente nel modello di Aronoff dove gli affissi vengono incorporati all'interno del componente delle RFP. Questa collocazione comporta anche una differente natura delle parole rispetto agli affissi:

Mentre le prime sono associate ad informazioni “categoriali” (ad es. ragazzo è “un nome”), i secondi sono associati ad informazioni “relazionali” (ad es. il suffisso *-bile* forma aggettivi da verbi). L'informazione $V \rightarrow A$, come si può vedere, è una RFP, anche se in forma abbreviata. In altre parole, in questa proposta, la rappresentazione di un affisso è la RFP che aggiunge l'affisso alla propria base.³³

³³ Scalise 1984, 73.

3.3 Morphology by Itself

In *Morphology by Itself* (1994), Aronoff fornisce argomenti a sostegno della trattazione dei fenomeni morfologici in un componente autonomo. Nel capitolo due di *Morphology by Itself*, Aronoff esamina il ruolo delle basi morfologiche nella lingua latina. Aronoff inizia discutendo il lavoro di Matthews (1962), precursore della morfologia basata su lessemi. Matthews si sofferma sulla *Priscianics o Parasitic formation* nella quale “one member of a paradigm seems to be formed not on the lexical root of the paradigm but instead on the stem of another member paradigm³⁴”. Viene proposto come esempio il caso della forma del participio futuro attivo in latino, marcato dal prefisso *-ur* (e concordato in numero, genere, caso con il nome o il soggetto cui è associato).

(1) Lauda-re : infinito

Laudat- : participio perfetto

Laudat-ur- : participio futuro

Matthews sostiene che il participio futuro attivo derivi dal participio perfetto solo sul piano della forma, sottolineando le discrepanze semantiche e la funzione di rimpiazzo del suffisso *-ur*, il quale sostituisce il significato del suffisso *-t*. Aronoff ricerca evidenze che supportino la tesi di una parentela tra i due tempi del participio. Si è vista la discrepanza sul piano sintattico, in quanto il perfetto è una forma passiva mentre il futuro è una forma attiva. Una seconda questione è la non presenza del participio perfetto nei verbi intransitivi, i quali non hanno ovviamente forma passiva. Questo rende difficile sostenere che il participio futuro derivi dal participio perfetto. Inoltre, nella grammatica latina è presente anche il supino, formalmente identico al participio perfetto. Si suppone che il participio perfetto sia stato preferito al supino come forma base solo a causa della bassa frequenza del supino, poco utilizzato nel periodo classico. Ma questo non è un argomento valido sul piano teorico a giustificare la scelta del participio perfetto come base e l'esistenza di due tempi verbali da cui derivare il participio futuro costituisce un problema. Aronoff ipotizza quindi l'esistenza di una base che riunisca tutto ciò che è comune alle tre forme (part. Perfetto, part. Futuro, supino) e lo definisce *terza base*. La sua peculiarità rispetto alle basi individuate in precedenza è l'impossibilità di associarla

³⁴ Aronoff 1994, 31.

a caratteristiche semantiche o sintattiche, in quanto, come si è visto sopra, le sue forme derivate differiscono sotto tutti questi aspetti. Si tratta di una base puramente morfologica attinente, sul piano grammaticale, alla categoria della “morfologia per sé stessa”.³⁵

3.4 Eredità teorica del lavoro di Aronoff

In conclusione, si possono riassumere come segue i principali cambi nell’ impostazione teorica della linguistica dovuti al lavoro di Aronoff: innanzi tutto, la morfologia diviene la branca della linguistica che studia il meccanismo mentale chiamato a dar conto della formazione delle parole e della loro struttura interna. Il morfema è l’unità base nello studio della struttura interna delle parole e viene definito “*the smallest linguistic pieces with a grammatical function*” (Aronoff, Feudeman, 2004, 2) e le stesse parole sono l’unità base per le regole di formazione di parola. Il morfema era stato definito anche come una associazione tra suono e significato; come si è visto, Aronoff obietta a questa definizione l’osservazione che alcuni morfemi hanno forma discontinua e significato variabile (e a volte sono del tutto privi di significato).

In secondo luogo, al componente del lessico ed ai processi di generazione di parola viene riconosciuto un carattere peculiare sulla base di una evidente differenza con la sintassi: la non identità tra parole possibili e parole grammaticalmente corrette. Aronoff descrive così la reazione dei parlanti all’ascolto o lettura di una parola nuova:

When we see or hear them, they leap out at us, for the simple reason that we have probably never seen or heard them before. It is interesting that novel words do this to us, because novel sentences do not. When you hear a new sentence, you generally don’t realize that this is the first time that you’ve heard it. And you don’t say to yourself, “What a remarkable sentence,” unless it happens to be one from Proust or Joyce or some other verbal artist. Many people have made the observation before that morphology differs from syntax in this way.³⁶

Infine, la morfologia viene definitivamente accreditata come componente autonomo del linguaggio. L’esistenza di lingue, come il Vietnamita, dove sono assenti parole morfologicamente complesse, può indurre a pensare che la morfologia non sia un

³⁵ Ovvero come si potrebbe tradurre letteralmente il titolo del libro di Aronoff in esame.

³⁶ Aronoff, Feudeman 2004, 5.

componente necessario della grammatica. Se la morfologia non è presente in tutte le lingue allora non la si può considerare un componente autonomo, ma solo un insieme di fenomeni appartenenti a sintassi e fonologia non presenti in tutte le lingue. Aronoff, in particolare all'interno di *Morphology By Itself*, obietta che esistono aspetti della morfologia in nessun modo attribuibili alla sintassi o alla fonologia. Uno dei principali indizi dell'esistenza di un componente morfologico è l'esistenza nelle lingue di classi all'interno delle quali raggruppare le parole, come ad esempio le cinque declinazioni del latino. La natura di queste classi è unicamente morfologica e non possono essere spiegate attraverso la sintassi o la fonologia. Nel caso delle declinazioni latine, la loro natura di classi morfologiche è ulteriormente rafforzata se si osserva che i nomi latini sono allo stesso tempo raggruppati in classi sintattiche (genere e numero) e che le classi morfologiche e le classi sintattiche si incrociano tra di loro, lavorando su piani grammaticali differenti.

Un secondo indizio è l'esistenza di morfemi privi di alcun significato sintattico, semantico o fonologico (v. *supra*, § 3.3). In conclusione, l'esistenza di classi esclusivamente morfologiche, e di morfemi non trattabili al di fuori della morfologia, prova la necessità di un componente morfologico autonomo. Questo non esclude la possibilità della morfologia di interagire con gli altri componenti grammaticali ed anzi si riconosce al componente morfologico il ruolo di tramite tra sintassi, fonologia e semantica:

Of course, morphology, probably more than any other component of language, interacts with all the rest, but it still has properties of its own.³⁷

³⁷ Aronoff, Feudeman 2004, 10.

Capitolo 4: approcci alla morfologia computazionale

Attualmente, l'interesse verso gli aspetti teorici e applicativi della formazione di parole è ampiamente legato alla necessità di applicazioni per l'automazione dei processi semantici. Infatti, strette sono le relazioni tra i processi derivazionali e i significati delle parole e generalmente forme derivate mediante il medesimo formativo condividono anche caratteristiche semantiche. Questa tendenza è diffusa ma non assoluta e si è visto nei capitoli precedenti come l'esistenza di parole derivate secondo una prospettiva esclusivamente morfologica (come nel caso della terza base dei verbi latini) sia una delle prove alla base della "morfologia per sé stessa". Tuttavia, la possibilità di dedurre automaticamente il significato di una parola a partire dalla sua base lessicale e dai suoi eventuali elementi formativi rimane particolarmente proficua sul versante applicativo e uno studio accurato della morfologia e dei mutamenti semantici associati a determinati fenomeni morfologici rimane imprescindibile nel campo della linguistica computazionale.

A questo scopo si sono sviluppati modelli computazionali per la morfologia orientati non solamente alla generazione di tutte le forme possibili, ma anche alla loro analisi. In questo capitolo si analizzerà in particolare la *Two-Level Morphology*, o Morfologia a Due Livelli, uno dei primi modelli computazionali orientato all'analisi morfologica. Si può osservare come l'approccio analitico derivi dalla necessità di inferire mutamenti semantici attraverso la morfologia e come questa tendenza sembrerebbe indirizzata verso una diminuzione dell'indipendenza del componente morfologico. Senza riaprire il tema relativo all'autonomia e all'importanza della morfologia, di cui si è già discusso nelle pagine precedenti, ci si limita a riportare la seguente affermazione ad opera di Johannesenn circa la *Two-Level Morphology*:

Koskenniemi's Two-level morphology (TM) has become well known since it was developed in 1983. One reason for this is probably that it is one of the few models within computational linguistics that has taken morphology seriously.³⁸

³⁸ Johannesenn 1990, 51.

4.1 La Linguistica Computazionale e i modelli teorici della Linguistica

La nascita e i primi sviluppi della disciplina denominata Linguistica Computazionale devono molto a Noam Chomsky e alla Grammatica Generativa: in particolare, furono l'approccio deduttivo e formalistico alla grammatica e alla linguistica (da parte dei generativisti) a fornire modelli e paradigmi per la linguistica computazionale. La distinzione tra linguaggi formali e linguaggi naturali viene demarcata da Noam Chomsky all'interno di *Syntactic Structures*, dove l'insieme dei linguaggi naturali, in contrapposizione a quelli formali, raccoglie tutti i linguaggi appresi naturalmente dall'uomo.³⁹

Un calcolatore è naturalmente orientato alla gestione di linguaggi formali e di conseguenza, la descrizione e il trattamento dei linguaggi naturali attraverso modelli formali è un passaggio essenziale per lo sviluppo della Linguistica Computazionale. Allo stesso modo l'idea di descrivere le regole formali in grado di generare o riconoscere frasi complesse risulta naturalmente proficua in un'ottica computazionale, dove un calcolatore esegue operazioni secondo un set di regole predeterminato. In seguito, la ricerca nel campo della linguistica computazionale si distaccò dal filone della linguistica "pura", adottando modelli matematico-statistici lontani dall'approccio teoretico dei grammatici. In particolare, i sistemi di trattamento automatico del testo più all'avanguardia si avvalgono di algoritmi di *Machine Learning* e *Deep Learning*. Da un punto di vista teorico è interessante osservare come questi sistemi forniscano i migliori risultati in termini di efficienza e applicabilità⁴⁰, senza che alla base vi sia una reale riflessione sulla natura del linguaggio. Ciò nonostante, i punti di contatto tra le due discipline rimangono numerosi e spesso le questioni di fondo e le criticità con cui si sono dovuti confrontare i ricercatori nei rispettivi campi sono state le stesse. A riprova di ciò, si esamineranno i primi approcci nel campo computazionale alla morfologia e al lessico.

4.2 Le applicazioni della Morfologia Computazionale

In qualsiasi campo non prettamente teoretico, è imprescindibile il concetto di "applicazione" e l'elaborazione di un modello computazionale è sempre orientata alla

³⁹ Tuttavia, lo stesso Chomsky giunge alla conclusione che modelli computazionale, come ad esempio gli automi a stati finiti, siano inadeguati a descrivere processi come la generazione di frasi di una lingua naturale (Cf. Chomsky 1956, 115).

⁴⁰ Per esempio, si basano su queste tecnologie i moderni sistemi di traduzione automatica e gli assistenti vocali intelligenti.

risoluzione di un problema. Secondo questa prospettiva Ten Hacken sottolinea le differenti finalità dei modelli puramente teorici e di quelli computazionali:

Models of derivational morphology have been developed both in computational linguistics and in theoretical linguistics. In both domains, the aim is to express relations between lexical entries. In theoretical linguistics, these relations are meant to correlate with the organization of the mental lexicon. For computer models, the main concern is that they serve a system solving a particular problem of computational linguistics.⁴¹

Nel caso della morfologia computazionale i problemi da risolvere e quindi le possibili applicazioni, sono numerosi. Se ne indicano alcuni dei principali e più comuni secondo lo schema proposto in *Computational Morphology*⁴²:

- Corretta segmentazione in morfemi
- Conversione da grafema a fonema
- Correzione di spelling: spesso non è sufficiente comparare l'input a una lista di parole per individuare gli errori. Una lista particolarmente estesa potrebbe includere parole rare con le quali potrebbero combaciare eventuali refusi, risultando corretti.
- *Stemming*: usato in particolare nell'*Information Retrieval*, lo *stemming* consiste nel ridurre ogni parola presente in un testo ad una singola forma del suo *stem* (non necessariamente la sua forma base, come nel caso del lemma nelle parole latine)

Tre delle possibili applicazioni della morfologia computazionale tra le quattro elencate in *Computational Morphology*, coinvolgono anche il livello lessicale.⁴³ Le operazioni vengono quindi effettuate sulle parole. Per questo motivo la base imprescindibile per ogni modello deve essere il dizionario, ovvero la lista delle parole esistenti su cui è possibile effettuare operazioni. Una delle principali criticità per i sistemi che devono implementare un dizionario è il *gap* tra le parole presenti

⁴¹ Ten Hacken 2017, 1.

⁴² Trost 2003, 37-38.

⁴³ Solo la conversione da grafema a fonema non coinvolge strettamente il componente lessicale.

in un dizionario e le parole esistenti di una lingua. Sul piano paradigmatico si possono adottare due differenti sistemi per la catalogazione delle parole. Il primo sistema è il cosiddetto *full-form lexicon*, dove tutte le forme possibili di tutte le parole vengono elencate e associate alla loro interpretazione (per “interpretazione” si intendono la divisione in morfemi e le caratteristiche morfosintattiche associate). Questo sistema ha come vantaggio la semplicità di analisi ma pecca dal punto di vista dell’efficienza, in quanto troppo ridondante e non adatto a gestire eventuali parole o forme non listate nell’elenco.

Il secondo sistema è quello del *lemma lexicon*, dove una forma canonica, definita “lemma”, viene utilizzata come forma rappresentativa di un intero paradigma. Il vantaggio è una maggiore efficienza del sistema ma a discapito della semplicità dei processi di analisi. Sul piano teorico si può osservare come venga privilegiata l’impostazione di Aronoff a scapito del Modello di Halle: i morfemi non sono alla base delle operazioni morfologiche ed ogni nuova forma viene generata a partire da una forma base pregressa. Sul piano pratico si può rilevare come il paradigma *full-form lexicon*, sia facilmente applicabile a lingue caratterizzate da una morfologia “povera”, come ad esempio l’Inglese, ma risulti inefficiente nel caso di lingue dalla morfologia più complessa, come il finlandese:

To store full wordforms, inflected and derivated, in the lexicon maybe possible for a language like English, with relatively poor morphology. But Koskeniemi saw that for Finnish, where a single verb can have between 12.000 and 18.000 different graphemic forms (included clitics), such a solution would not work. If the American computational linguists had been Red Indians speaking the Cherokee-language Oneida, instead of white and English-speaking, then they too would probably have developed a morphological model that could handle their verbs with up to 100.000 forms each.⁴⁴

4.3 Trasduttore a stati finiti

Si è visto in precedenza come nella morfologia generativa i processi di derivazione di parola si svolgano sequenzialmente tramite l’applicazione in successione ordinata di

⁴⁴ Johannesenn 1990, 51.

regole morfo-fonologiche e come ognuno di questi passaggi coinvolga un diverso livello di rappresentazione. Questa tipologia di approccio, efficiente dal punto di vista teorico per quanto concerne l'aspetto generativo, risulta inadeguata se applicata all'analisi: nel sistema generativo l'ordine dei passaggi influenza il risultato, rendendo il processo non reversibile. Lavorando su questo difetto applicativo della teoria morfologica generativa, nel 1972 C. Douglas Johnson, all'interno della dissertazione *Formal Aspects of Phonological Description*, ipotizza che le regole morfo-fonologiche possano essere descritte da trasduttori a stati finiti. Un trasduttore a stati finiti (FST) è una macchina a stati finiti (rispettando la definizione di macchina di Turing) la quale ha due nastri di registrazione: uno per l'input e uno per l'output. Questa è la principale differenza rispetto ad un automa a stati finiti (FSA), il quale possiede solamente un nastro di registrazione (per l'output).

Un FSA può essere utilizzato per definire un linguaggio formale specificando un set di stringhe accettate, mentre un FST definisce la relazione tra due differenti set di stringhe. Questa proprietà degli FST li rende particolarmente adatti al trattamento delle regole morfofonologiche, le quali posso essere considerate come relazioni tra forme profonde e forme di superficie.

Una proprietà matematica dei trasduttori è che ogni coppia di trasduttori in sequenza può essere sostituita da un singolo trasduttore equivalente. Ne consegue che ogni cascata di regole applicate in sequenza può essere descritta da un singolo trasduttore il quale mappa la forma lessicale in superficie e la forma lessicale profonda. Questa idea è illustrata nella figura 1.

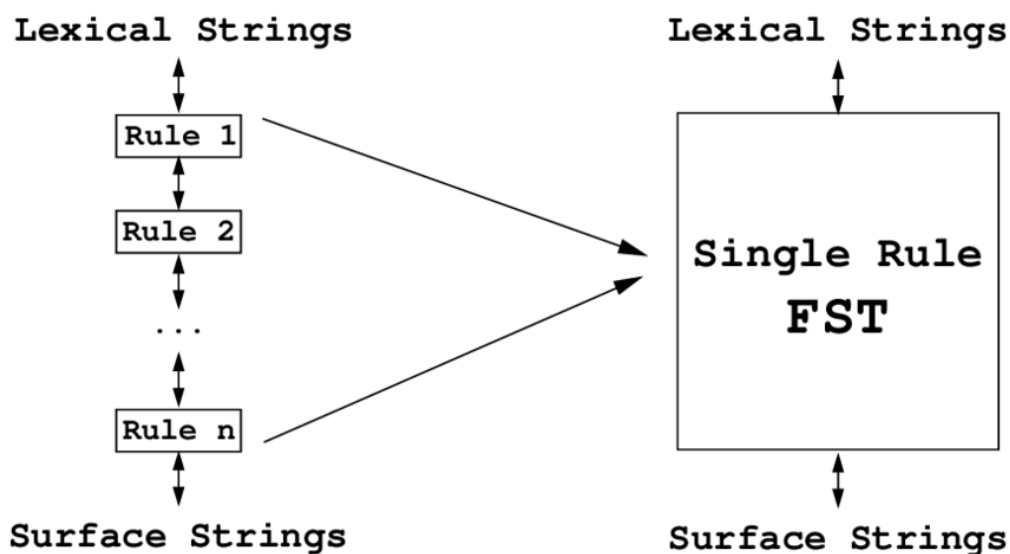


Figura 1: Trasduttori in sequenza e singolo trasduttore (Karttunen, Beesley 2001, 2).

L'intuizione di Douglas Johnson risolse solo parzialmente il problema dell'analisi, la quale risultava ancora molto più inefficiente della generazione. La complessità totale, anche riunendo le regole a cascata in un singolo trasduttore, rimaneva invariata. La medesima forma di superficie poteva essere generata in modi diversi e all'aumentare delle regole aumentava il numero di analisi possibili.

4.4 *Two - Level Morphology*

Nel 1980, due ricercatori della Xerox, Palo Alto, senza essere a conoscenza del lavoro di Douglas Johnson, riscoprono la possibilità di utilizzare gli FST per descrivere le regole morfofonologiche. Lo stesso anno, i due ricercatori della Xerox parteciparono a una conferenza sul *parsing* presso l'Università del Texas ad Austin ed incontrano Kimmo Koskenniemi un ricercatore finlandese interessato ai problemi computazionali dell'analisi morfologica. Koskenniemi era convinto che un'analisi morfologica efficiente non potesse essere effettuata con regole generative, anche se compilate attraverso FST. Al suo ritorno in Finlandia Koskenniemi inventò un sistema efficiente per la descrizione delle regole morfofonologiche e lo chiamò *Two-Level Morphology*, ovvero Morfologia a Due Livelli. Nel sistema di Koskenniemi i mutamenti morfologici non vengono descritti come una successione a cascata di regole con stati intermedi, bensì come *statements*, ovvero "affermazioni" le quali vincolano direttamente le realizzazioni lessicali alle corrispettive forme profonde. Le regole non vengono applicate in modo sequenziale ma in parallelo e

ogni regola vincola una certa corrispondenza tra forma profonda / forma di superficie e l'ambiente in cui la corrispondenza è consentita, richiesta o vietata.⁴⁵ I principi alla base della morfologia a due livelli sono:

- Le regole sono vincoli da simbolo a simbolo applicati in parallelo, non sequenzialmente come le regole di riscrittura.
- I vincoli possono riferirsi al contesto lessicale, al contesto di superficie, o ad entrambi i contesti allo stesso tempo.
- La ricerca lessicale e l'analisi morfologica vengono eseguite in tandem.

Le regole a “due livelli” sono bidirezionali e descrivono relazioni tra forme. Siccome i vincoli sono stabiliti da simbolo a simbolo, se una regola prevede un’aggiunta grafica o fonologica, il simbolo “0” viene utilizzato come corrispettivo nella forma profonda. Nella figura 2 vengono illustrati i vincoli che regolano l’alternanza *y - ie* nel plurale inglese:



Figura 2: rappresentazione a due livelli dell’alternanza *y\ ie* nel plurale inglese
(Karttunen, Beesley 2001, 5).

La sola applicazione delle regole in parallelo non è sufficiente a garantire una maggiore efficienza computazionale. Come nei sistemi sequenziali, una medesima forma può essere analizzata in modi diversi, aumentando il peso complessivo del processo. Tuttavia, l’esistenza di due soli livelli rende molto più facile il controllo lessicale. Nel sistema a due livelli la ricerca lessicale e l’analisi della forma di superficie vengono eseguite in tandem ed i possibili simboli in uscita vengono vincolati consultando il dizionario, il quale funge da filtro lessicale continuo.

⁴⁵ Cf. Karttunen, Beesley 2001, 4-5.

Koskenniemi's two-level morphology was the first practical general model in the history of computational linguistics for the analysis of morphologically complex languages. The language-specific components, the rules and the lexicon, were combined with a universal runtime engine applicable to all languages. The original implementation was primarily intended for analysis, but the model was in principle bidirectional and could be used for generation.⁴⁶

La Morfologia a Due Livelli venne da subito accettata e adottata nel campo della linguistica computazionale, tuttavia, le intuizioni linguistiche alla sua base non furono immediatamente riprese all'interno della linguistica teorica: molti argomenti, basati sull'assunzione a priori che una regola può riferirsi solo al contesto di input, furono avanzati a sostegno delle regole di riscrittura sequenziali. Il sistema dei vincoli tra simboli verrà ripreso soltanto nel quadro della *Optimality Theory*, teoria sviluppata proprio nel quadro di una critica alla tradizione "serialista" precedente.⁴⁷

⁴⁶ Cf. *ivi*, 4-5.

⁴⁷ Cf. *ivi*, 8.

Conclusioni

Nell'arco di tempo che va dal 1957 alla fine degli anni '90 i componenti lessicale e morfologico si sono sviluppati nel contesto della Grammatica Generativa. L'interesse verso questi due componenti è aumentato nel corso degli anni a partire da una situazione di partenza diametralmente opposta a quella conclusiva: i problemi morfologici e lessicali, trattati nella prima GGT nei componenti sintattico e fonologico, hanno guadagnato uno status di indipendenza rispetto agli altri componenti della grammatica. Né è derivata l'esigenza di un componente autonomo in grado di descrivere i meccanismi di formazione di parola e l'organizzazione del lessico mentale del parlante. Il primo modello di componente morfologico per la generazione delle parole è stato ipotizzato da Halle. Successivamente Aronoff ha corretto e riveduto tale modello, restringendo alle sole parole (e quindi escludendo i morfemi) il ruolo di formativi. Lo stesso Aronoff ha sistematizzato la teoria morfologica generativa, sostenendo l'esistenza di una morfologia 'per sé stessa' e dimostrando estesamente l'esistenza di fenomeni strettamente morfologici.

La medesima esigenza di un modello morfologico per la generazione delle parole è sorta nel campo della linguistica computazionale con finalità diverse, ovvero mettere a disposizione dei sistemi di trattamento automatico del linguaggio una lista di parole il più completa possibile, comparabile al lessico di un parlante di una determinata lingua. A questo scopo, si sono progettati modelli computazionali atti alla elaborazione della lista di parole di un linguaggio a partire da una lista di formativi e da un set di regole morfofonologiche. L'inferenza di tutte le parole esistenti di una lingua a partire da una lista di formativi è un problema complesso a causa della distanza esistente tra le nozioni di 'parola possibile' (ovvero generata applicando una regola generale a una base lessicale) e 'parola esistente'. La linguistica computazionale ha parzialmente mutuato i modelli generativi, ma allo stesso tempo si è scontrata con i limiti oggettivi di tali modelli sul versante dell'analisi. Le esigenze di un modello morfologico bidirezionale, funzionale all'analisi e alla generazione e performante da un punto di vista computazionale si incontrano nel modello della *Two-Level Morphology* di Kimmo Koskeniemi, presentato nel 1982. A riprova che le domande e le problematiche alla base di modelli

morfologici teorici e computazionali siano fondamentalmente le stesse, il modello di Koskeenniemi, nonostante l'approccio e le finalità prettamente applicative, contiene intuizioni (quali i vincoli da segno a segno) innovative non solo in un quadro strettamente computazionale, ma anche in una più estesa prospettiva linguistica teorica.

Bibliografia

ARONOFF 1976 = M. ARONOFF, *Word Formation in Generative Grammar*, Cambridge, MIT-press, 1976.

ARONOFF 1994 = M. ARONOFF, *Morphology by Itself: Stem and Inflectional Classes*, Cambridge, MIT-press, 1994.

ARONOFF FUDEMAN 2005 = M. ARONOFF, K. FUDEMAN, *What is Morphology?*, Oxford, Blackwell Publishing, 2005.

CHOMSKY = N. CHOMSKY, *Three Models for the Description of Language*, Cambridge, Department of Modern Languages and Research Laboratory of Electronics Massachusetts Institute of Technology, 1956.

TEN HACKEN 2017 = P. TEN HACKEN, *Computer Models and Mental Models of Derivational Morphology*, in E. LITTA, M. PASSAROTTI (eds.), *Proceedings of the Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, Milano, EDUCatt, 2017, 1-10.

HALLE 1973 = M. HALLE, *Prologomena to a Theory of Word Formation*, «Linguistic Inquiry» vol.4 n.1 (1973), 3-16.

JOHANNESSEN 1990 = J. B. JOHANNESSEN, *Is Two-Level Morphology a Morphological Model?*, in J. Pind, E. Rögnvaldsson (eds.), *Proceedings of the 7th Nordic Conference of Computational Linguistics (NODALIDA)*, Reykjavík, Institute of Lexicography, 1989, 51-60.

KARTUNNEN, BEESLEY 2001 = L. KARTUNNEN, K. R. BEESLEY, *A Short History of Two-Level Morphology*, Xerox Palo Alto Research Center, Xerox Research Centre Europe, 2001.

KOSKENNIEMI 1983 = K. KOSKENNIEMI, *Two level Morphology: A general Computational Model for Word-Form Recognition and Production*, University of Helsinki, Department of General Linguistic Publications N°11, 1983.

MEHRYAR 2004 = MEHRYAR MOHRI, *Weighted Finite-State Transducer Algorithms: An Overview*, Florham Park, AT&T Labs - Research Shannon Laboratory, 2004.

SCALISE 1990 = S. SCALISE, *Morfologia e Lessico*, Bologna, Il Mulino, 1990.

SPENCER 1991 = A. SPENCER, *Morphological Theory*, Oxford, Blackwell Publishing, 1991.

TROST = H. TROST, *Computational Morphology*, in R. MITKOV (ed.), *The Oxford Handbook of Computational Linguistics*, Oxford, Oxford University Press, 2003, 38-47.