

Il piano per un po'

In questa lezione e nella prossima, ci allontaneremo per un po' dal tema centrale di questo modulo, e cioè dallo studio delle capacità linguistiche degli animali.

Concentreremo invece la nostra attenzione su alcuni dispositivi astratti che sono stati impiegati per studiare le lingue naturali degli esseri umani:

- le grammatiche formali,
- gli automi.

Come vedremo in seguito, questi dispositivi sono utili anche per indagare le capacità linguistiche degli animali.

Un testo

Considerate questo testo:

Il sarpione zanzava la pirozzola. Il fibrido saluzzava, zulutando la cunzola. Un'esorzione scarzò la luftola. I flianti filibustavano lazzamente, reconando che i boni ciappi non vortessero zilcio.

Un altro testo

Ora considerate questo testo:

Gli sarpione zanzavi gli pirozzola. Fibrido saluzzava il, la zulutando cunzola. Un'esorzione la scarzò luftola. I flianti filibustava lazzamente, che i reconando boni zilcio ciappi non vortesse.

Livelli di accettabilità dei testi

I testi sono ambedue senza senso.

Ma c'è una differenza tra loro.

Il secondo testo è peggiore del primo, non vi pare?

Analisi delle differenze

Il sarpione zanzava la pirozzola. Il fibrido saluzzava, zulutando la cunzola. Un'esorzione scarzò la luftola. I flianti filibustavano lazzamente, reconando che i boni ciappi non vortessero zilcio.

Gli sarpione zanzavi gli pirozzola. Fibrido saluzzava il, la zulutando cunzola. Un'esorzione la scarzò luftola. I flianti filibustava lazzamente, che i reconando boni zilcio ciappi non vortesse.

La differenza tra i due testi si può spiegare così. Il primo testo è senza senso, ma l'ordine delle parole è accettabile in italiano: per esempio, l'articolo precede il nome. Inoltre, il soggetto si accorda con il predicato come in italiano: il soggetto di terza persona singolare *il sarpione* regge il verbo di terza persona singolare *zanzava*, il soggetto di terza persona singolare *il fibrido* regge il verbo di terza persona singolare *saluzzava*, ecc.

Il secondo testo è senza senso. Inoltre, l'ordine delle parole è completamente inaccettabile in italiano: per esempio, l'articolo *il* nella seconda frase compare da solo dopo il verbo. E il soggetto non si accorda con il predicato come in italiano: *gli sarpione zanzavi*, *I flianti filibustava*, ecc.

Benformatezza sintattica

Il sarpione zanzava la pirozzola. Il fibrido saluzzava, zulutando la cunzola. Un'esorzione scarzò la luftola. I flianti filibustavano lazzamente, reconando che i boni ciappi non vortessero zilcio.

Gli sarpione zanzavi gli pirozzola. Fibrido saluzzava il, la zulutando cunzola. Un'esorzione la scarzò luftola. I flianti filibustava lazzamente, che i reconando boni zilcio ciappi non vortesse.

La maggiore accettabilità del primo testo non ha dunque che fare con il significato (dal momento che tutti e due i testi ne sono privi).

La maggiore accettabilità del primo testo ha che fare con la nostra percezione che l'ordine delle parole e l'accordo tra soggetto e predicato siano corretti (in quanto sono come in italiano).

Quali siano gli ordini delle parole ammessi da una lingua e quale sia un corretto accordo tra il soggetto e il predicato dipende dalle regole *sintattiche* della lingua.

Il fatto che il secondo testo sembra peggiore del primo dimostra dunque che, in quanto parlanti dell'italiano, noi percepiamo un livello di correttezza (o benformatezza) che è indipendente dal significato e che dipende invece dal fatto che si osservino le regole sintattiche della lingua.

Idee verdi

In *Syntactic structures* (1957), N. Chomsky ha usato l'esempio seguente per sostenere che esiste un livello di benformatezza che è indipendente dal significato:

- (1) Green colorless ideas sleep furiously
 "Idee verdi senza colore dormono furiosamente"

Chomsky ha osservato che i parlanti percepiscono questo enunciato come grammaticale benché sia privo di senso. Questo perché l'enunciato è sintatticamente ben formato, cioè osserva le regole della sintassi dell'inglese.

Jabberwocky

L'esistenza di un un livello di benformatezza che è indipendente dal significato è stato sfruttato da Lewis Carroll nella poesia *Jabberwocky* in *Alice attraverso lo specchio*.

Questa poesia contiene la stanza seguente:

"'Twas brillig, and the slithy toves
 Did gyre and gimble in the wabe:
 All mimsy were the borogoves,
 And the mome raths outgrabe."

Questi versi sono privi di significato perché le parole *brillig*, *slithy*, *toves*, *gyre*, *gimble*, *wabe*, *mimsy*, *borogoves*, *mome*, *raths*, *outgrabe* sono inventate. Ma i parlanti inglesi percepiscono il testo come grammaticale.

Conoscenza del linguaggio

Quando conosciamo una lingua siamo dunque in grado riconoscere, tra le altre cose, che certe sequenze di parole sono sintatticamente ben formate in quella lingua, cioè sono *frasi* di quella lingua, e che altre sequenze di parole invece non lo sono.

Per esempio, sappiamo l'italiano e, per questo, siamo in grado di riconoscere che (2) è una frase dell'italiano mentre (3) non lo è:

(2) il principe bacia il rospo

(3) principe il rospo bacia il

Una domanda

- Come si spiega questa capacità di riconoscere quali sequenze di parole sono frasi dell'italiano e quali invece no?

Una risposta sbagliata

Qualcuno potrebbe pensare che la capacità di distinguere le sequenze di parole che sono frasi dell'italiano da quelle che non lo sono si possa spiegare così:

I parlanti dell'italiano hanno imparato una *lista* di frasi, (le frasi dell'italiano, appunto). Quando sentono una sequenza di parole, controllano se è nella lista che conoscono. Se c'è, allora quella sequenza è un frase dell'italiano, altrimenti no.

In realtà, questa spiegazione non funziona. Perché no?

Perché è sbagliata

Una ragione

La spiegazione che fa dipendere la capacità di riconoscere le frasi della lingua dalla conoscenza di una lista di frasi è sbagliata per almeno due ragioni.

Una è che un parlante competente dell'italiano è in grado di riconoscere come frasi delle sequenze di parole che non ha mai visto prima. Per esempio, è improbabile che voi abbiate sentito prima la sequenza di parole in (4), eppure siete in grado di riconoscere che è una frase dell'italiano:

(4) Guarda l'ippopotamo guercio a pallini blu.

Perché è sbagliata

Un'altra ragione

Un'altra ragione per cui la teoria della lista è sbagliata è che il numero di frasi dell'italiano, come abbiamo visto, è infinito:

- (5)
- a. Terry sa che Armando è partito
 - b. Armando sa che Terry sa che Armando è partito
 - c. Terry sa che Armando sa che Terry sa che Armando è partito
 - d. ...

Dunque, la conoscenza attraverso la quale i parlanti dell'italiano sono in grado di giudicare se una sequenza di parole è una frase dell'italiano oppure no non può consistere nel conoscere una lista di frasi, dal momento che questa lista, essendo infinita, non può essere immagazzinata nel nostro cervello, che ha una capacità finita.

Grammatiche

Una spiegazione migliore della nostra capacità di riconoscere se una sequenza di parole è una frase dell'italiano oppure no è che la nostra conoscenza dell'italiano consiste nella conoscenza di un *insieme finito di regole* che, tra le altre cose, ci consentono appunto di riconoscere le frasi dell'italiano.

L'insieme di regole che i parlanti di una lingua conoscono e per mezzo delle quali sono in grado di parlare e comprendere la lingua formano la *grammatica* della lingua.

Teoria delle grammatiche formali

Una domanda che gli studiosi delle lingue naturali umane si pongono è questa: come sono fatte le grammatiche di queste lingue? Che forma hanno le regole di queste grammatiche?

Per il momento, restringeremo la nostra attenzione agli aspetti sintattici delle grammatiche, vale a dire ci concentreremo sulle regole che permettono ai parlanti di riconoscere se una certa sequenza di parole è una frase oppure no. La domanda che ci poniamo è dunque questa:

- qual è la *forma* delle regole attraverso le quali gli esseri umani sono in grado di riconoscere le frasi della loro lingua?

Uno strumento per indagare la questione è la teoria delle grammatiche formali. Questa teoria studia la relazione tra diversi tipi di regole sintattiche e i linguaggi che è possibile generare attraverso queste regole.

Alcuni elementi della teoria

- Nelle pagine seguenti, presenterò alcune nozioni di base della teoria delle grammatiche formali.
- In primo luogo, introdurrò una definizione delle nozioni di grammatica e di linguaggio.
- Poi, descriverò delle grammatiche di tipi diversi ed esaminerò la relazione tra queste grammatiche e i linguaggi che esse generano.

Alfabeti e stringhe

- Un *alfabeto* è un insieme finito di simboli;
- una *stringa* su un alfabeto è una sequenza finita (eventualmente vuota) di occorrenze di simboli dell'alfabeto;
- una stringa senza simboli è detta *la stringa vuota* ed è denotata da ϵ ;
- La lunghezza di una stringa ψ è il numero delle occorrenze di simboli in ψ ;
- la totalità di tutte le stringhe possibili, inclusa la stringa vuota, su un alfabeto V è denotata da V^* .

Esempi di alfabeto

- L'alfabeto latino $\{a, b, c, \dots, z\}$ è un alfabeto secondo la nostra definizione.
- L'insieme costituito dai simboli 0 e 1 è un alfabeto secondo la nostra definizione.
- Il lessico dell'italiano (l'insieme delle parole dell'italiano) è un alfabeto secondo la nostra definizione.
- Da un punto di vista formale, ogni insieme finito può essere un alfabeto.

Esempi di stringhe

- La sequenza *melone* è una stringa sull'alfabeto latino.
- Le sequenze seguenti sono stringhe sull'alfabeto latino: *ab*, *bc*, *abc*, *abbb*.
- La sequenza *0111011* è una stringa sull'alfabeto $\{0, 1\}$.
- Le stringhe *ab* e *bc* sull'alfabeto latino sono di lunghezza 2. La stringa *abc* sull'alfabeto latino è di lunghezza 3. La stringa *abbb* sull'alfabeto latino è di lunghezza 4. La stringa *melone* sull'alfabeto latino è di lunghezza 6. La stringa *0111011* sull'alfabeto $\{0, 1\}$ è di lunghezza 7.

Linguaggi

- Qualsiasi insieme di stringhe su un alfabeto è un *linguaggio* (su quell'alfabeto).
- Dal momento che V^* è la totalità di tutte le stringhe su un alfabeto V , secondo la definizione precedente, qualsiasi sottoinsieme di V^* è un linguaggio.
- Se un linguaggio è un insieme finito di stringhe, possiamo descriverlo elencando tutte le sue stringhe. Per esempio, $\{aba, czr, d, f\}$ è un linguaggio sull'alfabeto $\{a, b, c, \dots, z\}$.
- Se un linguaggio è un insieme infinito di stringhe, tuttavia, non possiamo descriverlo elencando tutte le sue stringhe. Lo descriviamo invece per mezzo di una grammatica.

Grammatiche formali

Una *grammatica formale* (o semplicemente *grammatica*) è una quadrupla $G=(V_T, V_N, S, R)$, dove:

- V_T è un insieme finito di simboli terminali, o alfabeto terminale;
- V_N è un insieme finito di simboli non terminali, o alfabeto non terminale; (V_T e V_N non hanno simboli in comune).
- S è il simbolo iniziale, o simbolo di frase (un membro di V_N).
- R è un insieme di regole della forma $\psi \rightarrow \omega$, dove ψ e ω sono stringhe (su $V_T \cup V_N$) e ψ non consiste interamente di simboli terminali.

Un esempio di grammatica formale

$G_1=(V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{a, b\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A, B\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguente (uso le parentesi graffe per indicare insiemi):

$$\{S \rightarrow ABS, \\ S \rightarrow e, \\ AB \rightarrow BA, \\ BA \rightarrow AB, \\ A \rightarrow a, \\ B \rightarrow b\}$$

Applicazione delle regole

Le regole della grammatica funzionano così:

$\psi \rightarrow \omega$ significa che, ogniqualvolta ψ occorre come una sottostringa di una stringa data, quell'occorrenza di ψ può essere sostituita da ω per formare una nuova stringa.

Per esempio, possiamo usare la regola

$$S \rightarrow ABS$$

per formare la stringa ABABS a partire dalla stringa ABS.

Derivazioni

Data una grammatica $G=\{V_T, V_N, S, R\}$, una *derivazione* in G è una sequenza di stringhe x_1, x_2, \dots, x_n tale che

- x_1, x_2, \dots, x_n sono costituite da simboli in $V_T \cup V_N$,
- $x_1 = S$,
- ogni x_i (where $2 \leq i \leq n$) è ottenuta da x_{i-1} attraverso una applicazione di una regola in R .

Esempi di derivazioni

Consideriamo di nuovo la grammatica G_1 :

- V_T (l'alfabeto terminale): {a,b}
- V_N (l'alfabeto non terminale): {S,A,B}
- il simbolo iniziale: S
- R (l'insieme di regole):

{S \rightarrow ABS,
S \rightarrow e,
AB \rightarrow BA,
BA \rightarrow AB,
A \rightarrow a,
B \rightarrow b}

Le sequenze seguenti sono derivazioni in questa grammatica (potete infatti controllare che ogni stringa successiva a S è derivata dalla stringa precedente attraverso l'applicazione di una regola in R):

- S, ABS, ABABS, ABABABS
- S, ABS, ABABS, ABAB
- S, ABS, BAS, BAABS, BAAB

Notazione per le derivazioni

Le derivazioni in una grammatica possono anche essere rappresentate per mezzo di questa notazione (dove $\alpha \Rightarrow_G \beta$ sta per "la stringa β è ottenuta da α per mezzo dell'applicazione di una regola di G"):

$$x_1 \Rightarrow_G x_2 \Rightarrow_G \dots \Rightarrow_G x_n$$

Per esempio, le derivazioni seguenti in G_1

- S, ABS, ABABS, ABABABS
- S, ABS, ABABS, ABAB
- S, ABS, BAS, BAABS, BAAB

possono anche essere rappresentate così:

- $S \Rightarrow_{G_1} ABS \Rightarrow_{G_1} ABABS \Rightarrow_{G_1} ABABABS$
- $n S \Rightarrow_{G_1} ABS \Rightarrow_{G_1} ABABS \Rightarrow_{G_1} ABAB$
- $S \Rightarrow_{G_1} ABS \Rightarrow_{G_1} BAS \Rightarrow_{G_1} BAABS \Rightarrow_{G_1} BAAB$

Linguaggi generati da grammatiche

- Una grammatica G *genera* una stringa α sull'alfabeto terminale V_T se c'è una derivazione x_1, x_2, \dots, x_n in G tale che $x_n = \alpha$.
- L'insieme delle stringhe di simboli terminali generati da una grammatica G è *il linguaggio generato da G*.

Un esempio di stringa generata da una grammatica

La grammatica G_1 ci permette di formare le stringhe seguenti:

ABABABS
ABAB
BAAB
abab

Tuttavia solo 'abab' è una stringa del linguaggio generato da G_1 , in quanto solo 'abab', tra queste stringhe, è una stringa di simboli terminali generata dalle regole di G_1 .

Un esempio di linguaggio generato da una grammatica

Si può mostrare che il linguaggio generato da G_1 è l'insieme delle stringhe sull'alfabeto $\{a,b\}$ che contengono lo stesso numero di a e di b.

$G_1 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale): $\{a,b\}$;
- V_N (l'alfabeto non terminale): $\{S,A,B\}$;
- il simbolo iniziale: S;
- R (l'insieme di regole):

$$\{S \rightarrow ABS, \\ S \rightarrow e, \\ AB \rightarrow BA, \\ BA \rightarrow AB, \\ A \rightarrow a, \\ B \rightarrow b\}$$

Un'altra grammatica

Ecco un'altra grammatica che, a differenza della precedente, usa un insieme di parole italiane come alfabeto terminale:

$G_2 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{\text{Gianni, buon, rosso, vino, beve}\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, NP, AP, VP, N, A, V\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguente:

$$\{S \rightarrow NP VP, \\ VP \rightarrow V NP, \\ NP \rightarrow AP NP, \\ NP \rightarrow N, \\ AP \rightarrow A, \\ A \rightarrow \text{buon}, \\ A \rightarrow \text{rosso}, \\ N \rightarrow \text{Gianni}, \\ N \rightarrow \text{vino}, \\ V \rightarrow \text{beve}\}$$

Esercizio

Domanda

Le stringhe seguenti appartengono tutte al linguaggio generato da G_2 :

- (6) a. Gianni beve vino
b. Gianni beve buon vino
c. buon vino beve Gianni

La grammatica G_2 genera inoltre le stringhe seguenti:

- (7) a. rosso Gianni beve rosso buon Gianni
b. buon vino beve rosso rosso buon rosso buon vino

Come si derivano queste stringhe?

Esercizio

Risposta

Ecco le derivazioni in G_2 per le stringhe (6)a-(7)a (le altre derivazioni sono lasciate come esercizio):

$S \rightarrow NP VP$	$VP \rightarrow V NP$
$NP \rightarrow AP NP$	$NP \rightarrow N$
$AP \rightarrow A$	$A \rightarrow \text{buon}$
$A \rightarrow \text{rosso}$	$N \rightarrow \text{Gianni}$
$N \rightarrow \text{vino}$	$V \rightarrow \text{beve}$

$S \Rightarrow_{G_2} NP VP \Rightarrow_{G_2} NP V NP \Rightarrow_{G_2} N V NP \Rightarrow_{G_2} \text{Gianni} V NP \Rightarrow_{G_2} \text{Gianni} \\ \text{beve} NP \Rightarrow_{G_2} \text{Gianni} \text{ beve} N \Rightarrow_{G_2} \text{Gianni} \text{ beve} \text{ vino}$

$S \Rightarrow_{G_2} NP VP \Rightarrow_{G_2} NP V NP \Rightarrow_{G_2} AP NP V NP \Rightarrow_{G_2} AP N V NP \Rightarrow_{G_2} A \\ N V NP \Rightarrow_{G_2} A N V AP NP \Rightarrow_{G_2} A N V AP AP NP \Rightarrow_{G_2} A N V A AP NP \Rightarrow_{G_2} \\ A N V A A NP \Rightarrow_{G_2} A N V A A N \Rightarrow_{G_2} \text{rosso} N V A A N \Rightarrow_{G_2} \text{rosso} \text{ Gianni} \\ V A A N \Rightarrow_{G_2} \text{rosso} \text{ Gianni} \text{ beve} A A N \Rightarrow_{G_2} \text{rosso} \text{ Gianni} \text{ beve} \text{ rosso} A N \\ \Rightarrow_{G_2} \text{rosso} \text{ Gianni} \text{ beve} \text{ rosso} \text{ buon} N \Rightarrow_{G_2} \text{rosso} \text{ Gianni} \text{ beve} \text{ rosso} \text{ buon} \\ \text{Gianni}$

Grammatiche di tipo 0

La definizione di grammatica formale che abbiamo assunto non impone alcuna restrizione sulla forma delle regole della grammatica, eccetto per la condizione che la stringa a sinistra della freccia non deve consistere interamente di simboli terminali.

Si ricordi infatti che una *grammatica formale* è qualunque quadrupla $G=(V_T, V_N, S, R)$, dove:

- V_T è un insieme finito di simboli terminali, o alfabeto terminale;
- V_N è un insieme finito di simboli non terminali, o alfabeto non terminale; (V_T e V_N non hanno simboli in comune).
- S è il simbolo iniziale, o simbolo di frase (un membro di V_N).
- R è un insieme di regole della forma $\psi \rightarrow \omega$, dove ψ e ω sono stringhe (su $V_T \cup V_N$) e ψ non consiste interamente di simboli terminali.

Queste grammatiche sono dette anche *grammatiche di tipo 0* o *grammatiche non ristrette*.

Altri tipi di grammatiche

Imponendo condizioni ulteriori sulla forma delle regole, possiamo ottenere grammatiche di tipi particolari. Ecco i nomi di alcuni tipi di grammatiche che sono state studiate:

- grammatiche di tipo 1 (grammatiche *context-sensitive*)
- grammatiche di tipo 2 (grammatiche *context-free*)
- grammatiche di tipo 3 (grammatiche regolari)

Grammatiche di tipo 1 (context-sensitive)

Una grammatica di tipo 1 (o grammatica *context-sensitive*) è una grammatica formale le cui regole sono della forma

$$\varphi A \psi \rightarrow \varphi \omega \psi,$$

dove A è un simbolo non terminale, φ e ψ sono stringhe arbitrarie (eventualmente vuote) e ω è una stringa arbitraria non vuota.

Le grammatiche di questo tipo sono dette anche *context-sensitive*, perché la possibilità di sostituire un simbolo non terminale in una stringa dipende dai simboli adiacenti (cioè, dipende dal contesto in cui il simbolo non terminale si trova).

Una domanda

La grammatica G_3 è *context-sensitive*?

$G_3=(V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{a, b, c\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A, B, C, X, Y\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguente:

- (i) $S \rightarrow A$
- (ii) $A \rightarrow aABC$
- (iii) $A \rightarrow abC$
- (iv) $bB \rightarrow bb$
- (v) $bC \rightarrow bc$
- (vi) $cC \rightarrow cc$
- (vii) $CB \rightarrow XB$
- (viii) $XB \rightarrow XY$
- (ix) $XY \rightarrow BY$
- (x) $BY \rightarrow BC$

La risposta

La grammatica G_3 è *context-sensitive*, in quanto ogni regola in (i-x) è della forma $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ e ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

$G_3 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme {a, b, c}
- V_N (l'alfabeto non terminale) è l'insieme {S, A, B, C, X, Y}
- S è il simbolo iniziale
- R è l'insieme di regole seguente:

- (i) $S \rightarrow A$
- (ii) $A \rightarrow aABC$
- (iii) $A \rightarrow abC$
- (iv) $bB \rightarrow bb$
- (v) $bC \rightarrow bc$
- (vi) $cC \rightarrow cc$
- (vii) $CB \rightarrow XB$
- (viii) $XB \rightarrow XY$
- (ix) $XY \rightarrow BY$
- (x) $BY \rightarrow BC$

Verifichiamo!

Verifichiamo per ciascuna regola in (i-x) che essa ha la forma richiesta dalla definizione di grammatica *context-sensitive*.

La forma delle regole**Le regole (i-iii)**

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Le regole (i-iii) di G_3 sono di questa forma con φ and ψ vuote.

- (i) $S \rightarrow A$
- (ii) $A \rightarrow aABC$
- (iii) $A \rightarrow abC$

La forma delle regole**La regola (iv)**

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Nella regola (iv), b è un'istanza di φ , B è un'istanza di A , b è un'istanza di ω e ψ è vuota.

- (iv) $bB \rightarrow bb$

La forma delle regole

La regola (v)

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Nella regola (v), b è un'istanza di φ , C è un'istanza di A , c è un'istanza di ω e ψ è vuota.

$$(v) \quad bC \rightarrow bc$$

La forma delle regole

La regola (vi)

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Nella regola (vi), c è un'istanza di φ , C è un'istanza di A , c è un'istanza di ω e ψ è vuota.

$$(vi) \quad cC \rightarrow cc$$

La forma delle regole

La regola (vii)

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Nella regola (vii), φ è vuota, B è un'istanza di ψ , C è un'istanza di A e X è un'istanza di ω .

$$(vii) \quad CB \rightarrow XB$$

La forma delle regole

La regola (viii)

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Nella regola (viii), ψ è vuota, X è un'istanza di φ , B è un'istanza di A e Y è un'istanza di ω .

$$(viii) \quad XB \rightarrow XY$$

La forma delle regole

La regola (ix)

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Nella regola (ix), φ è vuota, Y è un'istanza di ψ , X è un'istanza di A e B è un'istanza di ω .

$$(ix) \quad XY \rightarrow BY$$

La forma delle regole

La regola (x)

In una grammatica *context-sensitive* le regole hanno questa forma: $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

Nella regola (x), ψ è vuota, B è un'istanza di φ , Y è un'istanza di A e C è un'istanza di ω .

$$(x) \quad BY \rightarrow BC$$

Una domanda

La grammatica G_4 è *context-sensitive*?

$G_4 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{a, b, c\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A, B, C, X, Y\}$
- S è il simbolo iniziale.
- R è l'insieme di regole seguente:

- (i) $S \rightarrow A$
- (ii) $A \rightarrow aABC$
- (iii) $A \rightarrow abC$
- (iv) $bB \rightarrow bb$
- (v) $bC \rightarrow b$
- (vi) $CB \rightarrow XB$
- (vii) $XB \rightarrow XY$
- (viii) $XY \rightarrow BY$
- (ix) $BY \rightarrow BC$

La risposta

La grammatica G_4 non è *context-sensitive* (anche se è una grammatica non ristretta) in quanto la regola (v) non è della forma $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ and ψ sono stringhe (eventualmente vuote) e ω è una stringa non vuota.

$$(v) \quad bC \rightarrow b$$

Infatti, dal momento che ω non può essere vuota, in (v) b deve essere un'istanza di ω . A sinistra della freccia, C deve essere un'istanza di A (in quanto C è l'unico simbolo non terminale) e b deve essere un'istanza di φ . Dunque, φ occorre a sinistra ma non a destra della freccia, come sarebbe invece richiesto dalla definizione di grammatica *context-sensitive*.

Grammatiche di tipo 2 (context-free)

Una grammatica di tipo 2 (o *context-free*) è una grammatica formale le cui regole sono della forma

$$A \rightarrow \omega,$$

dove A è un simbolo non terminale e ω è una stringa arbitraria.

Le grammatiche di questo genere sono chiamate *context-free* perché la possibilità di sostituire un simbolo non terminale in una stringa è indipendente dai simboli adiacenti (dunque, è indipendente dal contesto)

Un esempio di grammatica context-free

La grammatica G_5 è *context-free*:

$G_5 = (V_T, V_N, S, R)$, where

- V_T (l'alfabeto terminale) è l'insieme $\{a, b\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguenti:

$$(i) \quad S \rightarrow A$$

$$(ii) \quad A \rightarrow aAb$$

$$(iii) \quad A \rightarrow ab$$

Una domanda

La grammatica G_5 è *context-sensitive* (oltre a essere *context-free*)?

$G_5 = (V_T, V_N, S, R)$, where

- V_T (l'alfabeto terminale) è l'insieme $\{a, b\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguenti:

$$(i) \quad S \rightarrow A$$

$$(ii) \quad A \rightarrow aAb$$

$$(iii) \quad A \rightarrow ab$$

La risposta

La grammatica G_5 è *context-sensitive*. Infatti, le regole di G_5 sono casi di regole *context-sensitive* con φ and ψ vuote.

$G_5 = (V_T, V_N, S, R)$, where

- V_T (l'alfabeto terminale) è l'insieme $\{a, b\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguenti:

$$(i) \quad S \rightarrow A$$

$$(ii) \quad A \rightarrow aAb$$

$$(iii) \quad A \rightarrow ab$$

Un'altra domanda

La grammatica G_3 è *context-free* (oltre a essere *context-sensitive*)?

$G_3 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{a, b, c\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A, B, C, X, Y\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguente:

- (i) $S \rightarrow A$
- (ii) $A \rightarrow aABC$
- (iii) $A \rightarrow abC$
- (iv) $bB \rightarrow bb$
- (v) $bC \rightarrow bc$
- (vi) $cC \rightarrow cc$
- (vii) $CB \rightarrow XB$
- (viii) $XB \rightarrow XY$
- (ix) $XY \rightarrow BY$
- (x) $BY \rightarrow BC$

La risposta

La grammatica G_3 non è *context-free*, in quanto le regole delle grammatiche *context-free* devono avere solo un simbolo non terminale a sinistra della freccia.

$G_3 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{a, b, c\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A, B, C, X, Y\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguente:

- (i) $S \rightarrow A$
- (ii) $A \rightarrow aABC$
- (iii) $A \rightarrow abC$
- (iv) $bB \rightarrow bb$
- (v) $bC \rightarrow bc$
- (vi) $cC \rightarrow cc$
- (vii) $CB \rightarrow XB$
- (viii) $XB \rightarrow XY$
- (ix) $XY \rightarrow BY$
- (x) $BY \rightarrow BC$

Grammatiche di tipo 3 (regolari)

Una grammatica di tipo 3 (o regolare) è una grammatica formale le cui regole sono della forma in (a) o della forma in (b):

- a. $A \rightarrow xB$
- b. $A \rightarrow x,$

dove A e B sono simboli non-terminali e x è una stringa arbitraria di simboli terminali (x è in V_T^*)

Un esempio di grammatica regolare

La grammatica G_6 è regolare:

$G_6 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{0, 1\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A, B\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguente:

- (i) $S \rightarrow 1B$
- (ii) $S \rightarrow 1$
- (iii) $A \rightarrow 1B$
- (iv) $B \rightarrow 0A$
- (v) $A \rightarrow 1$

Una domanda

La grammatica G_7 è regolare?

$G_7 = (V_T, V_N, S, R)$, dove

- V_T (l'alfabeto terminale) è l'insieme $\{0, 1\}$
- V_N (l'alfabeto non terminale) è l'insieme $\{S, A, B\}$
- S è il simbolo iniziale
- R è l'insieme di regole seguente:

- (i) $S \rightarrow 1B$
- (ii) $S \rightarrow 1$
- (iii) $A \rightarrow 1AB$
- (iv) $B \rightarrow 0A$
- (v) $A \rightarrow 1$

La risposta

La grammatica G_7 non è regolare, in quanto la regola (iii) permette di riscrivere A con una stringa che contiene più di un simbolo non terminale.

$$(iii) \quad A \rightarrow 1AB$$

Sommario dei tipi di grammatica

Forma delle regole

- Grammatiche di tipo 0 (grammatiche non ristrette): $\psi \rightarrow \omega$, dove ψ e ω sono stringhe e ψ non consiste interamente di simboli terminali.
- Grammatiche di tipo 1 (grammatiche context-sensitive): $\varphi A \psi \rightarrow \varphi \omega \psi$, dove A è un simbolo non terminale, φ e ψ sono stringhe arbitrarie (eventualmente vuote) and ω è una stringa arbitraria non vuota.
- Grammatiche di tipo 2 (grammatiche context-free): $A \rightarrow \omega$, dove A è un simbolo non terminale e ω una stringa arbitraria.
- Grammatiche di tipo 3 (grammatiche regolari): $A \rightarrow xB$ o $A \rightarrow x$, dove A and B sono simboli non terminali e x è una stringa arbitraria in V_T^* .

Inclusione tra grammatiche

- Come abbiamo visto, alcune grammatiche di tipo 0 (non ristrette) non sono *context sensitive* (la grammatica G_4 , ad esempio). Ma tutte le grammatiche *context-sensitive* sono di tipo 0, in quanto la stringa a sinistra della freccia nelle grammatiche *context-sensitive* non consiste interamente di simboli terminali. Le grammatiche *context-sensitive* sono grammatiche di tipo 0 che non permettono regole della forma $\psi \rightarrow e$ e che permettono la sostituzione di un simbolo non terminale solo in un contesto fisso.
- Invece, secondo le nostre definizioni, non solo ci sono delle grammatiche *context sensitive* che non sono *context-free* (per esempio, la grammatica G_3); è anche vero che non tutte le grammatiche *context-free* sono *context-sensitive*, in quanto le grammatiche *context-free*, a differenza di quelle *context-sensitive*, permettono regole della forma $A \rightarrow e$.
- Tutte le grammatiche regolari sono *context-free* (dal momento che tutte le regole delle grammatiche regolari sostituiscono i nodi non terminali indipendentemente dal contesto). Ma non tutte le grammatiche *context-free* sono regolari (come mostra la grammatica G_7).

Una domanda

Considerate la grammatica G_9 (dove i simboli non terminali sono S e A , il simbolo iniziale è S e i simboli terminali sono $0, 1$):

- (i) $S \rightarrow A$
- (ii) $A \rightarrow A0A$
- (iii) $A \rightarrow 1$

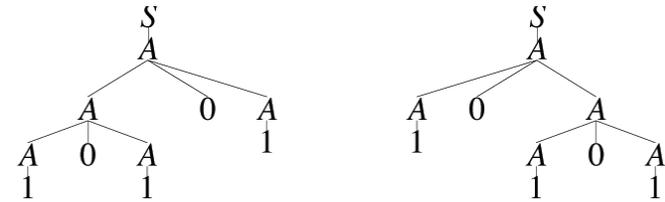
In questa grammatica, la stringa 10101 può essere derivata in più di un modo.

Come?

Ambiguità strutturale

- (i) $S \rightarrow A$
- (ii) $A \rightarrow A0A$
- (iii) $A \rightarrow 1$

Ecco due modi diversi di derivare 10101 :



In un caso di questo genere, in cui le due derivazioni danno luogo ad alberi diversi, si dice che la stringa 10101 è strutturalmente ambigua.