

Constituenți imediați

(engl. *immediate constituents* (IC), fr. *constituents immédiats*)

Un constituent imediat acea parte de enunț care este o unitate funcțională dotată cu sens și care face parte dintr-o unitate sintactică mai mare. Constituenții imediați formează o ierarhie, care pornește, în demersul analitic, de la enunț (frază sau propoziție), continuă cu sintagme, cu cuvinte și se oprește la nivelul morfemelor. Este o metodă care a jucat un rol important în dezvoltarea spectaculoasă a sintaxei în a doua jumătate a secolului trecut, ca și la ștergerea frontierei dintre morfologie și sintaxă (cu apariția morfosintaxei). Identificarea constituenților imediați bazată pe criterii formale riguroase, distribuționale, a fost importantă nu numai în lingvistică, dar și în domenii conexe, cum ar fi informatica (lingvistica computațională, traducerea automată) sau didactica limbilor străine.

Studiul constituenților imediați a început în cadrul distribuționalismului american, ca metodă de analiză morfosintactică. Într-o a doua etapă, Chomsky (1957) a creat o gramatică generativă de constituenți, care apoi a ajuns să facă parte din modelul generativ 'clasic', din *Aspects* (Chomsky 1965).

1. Analiza în constituenți imediați

Analiza în constituenți imediați (ACI) a apărut ca o metodă de descriere structurală care constă din descompunerea unei propoziții sau fraze în constituenți funcționali (părți) mai mici (v. SINTAGME), care vor fi la rândul lor descompuși în proprii constituenți, până la nivelul morfemelor (v. MORFEM). Analiza pune astfel în evidență structura ierarhică a enunțului.

Ca la orice analiză lingvistică, punctul de plecare al metodei ACI este observația evidentă că secvențele lingvistice nu sunt simple înlănțuiri (concatenări, serii) de elemente așa cum ar părea, pentru că forma elementelor lingvistice fiind sonoră, este percepută ca o succesiune liniară (linearitatea semnificativului, v. SEMN LINGVISTIC). Între elementele lingvistice există nu numai legături lineare, ci și raporturi structurale. Avem o secvență liniară dacă, în cazul succesiunii $a + b + c$, apariția elementului a determină apariția elementului b care, la rândul lui, determină apariția elementului c . Uneori relațiile lingvistice coincid cu ordinea liniară: într-o secvență ca *băiat inteligent*, elementul *băiat* (masculin singular) determină forma adjectivului *inteligent* (masculin singular). Această caracteristică nu se regăsește în secvența *acest băiat inteligent*, deoarece nu putem spune că ocurența demonstrativului *acest* determină apariția substantivului *băiat*. Cuvântul *băiat* este elementul lingvistic care determină forma elementelor *acest* și *inteligent*. În gramatica tradițională se spune că *băiat* este elementul principal, regent, iar cuvintele *acest* și *inteligent* sunt elemente care îi sunt subordonate. Gramatica structurală și cea generativă sunt interesate în primul rând de relațiile structurale.

Gramatica tradițională a propus, și ea, o analiză a enunțurilor, care prezintă două niveluri: (i) la primul nivel sunt identificate clasele de cuvinte care au un comportament lingvistic similar (substantiv, adjectiv, pronume, verb, adverb, prepoziție, conjuncție, interjecție); sunt identificate și descrise toate formele pe care le prezintă fiecare clasă (**cazul**, pentru substantiv, articol, adjectiv, articol, pronume, **timpul** pentru verb și unele adverbe, **persoană** pentru unele pronume și verbe, **numărul** pentru substantiv, adjectiv, articol, pronume și verb, etc.); acest studiu este numit 'morfologie'; (ii) la cel de al doilea nivel, se studiază relațiile care se stabilesc între diferitele cuvinte și formele lor morfologice, numite 'funcții': un substantiv sau pronume la nominativ îndeplinește de obicei funcția de subiect sau de nume predicativ, formele de acuzativ exprimă de obicei funcția de complement direct, adjectivele calificativ (ca *mare*) poate fi determinat numai de un adverb (ca *foarte*, *mai*), etc.; această parte a gramaticii se numește 'sintaxă'.

În prima jumătate a secolului al XX^{lea} mulți lingviști, mai ales din Statele Unite, au considerat că analiza propusă de gramatica tradițională nu este suficient de riguroasă și nici îndeajuns de precisă. De exemplu, încă din gramaticile grecești sau latinești ale antichității, unele clase morfologice sunt definite semantic (substantiv, adjectiv, verb, adverb), altele sunt definite funcțional (pronume, conjuncție, prepoziție). După cum se știe, substantivele sunt definite ca fiind clasa de cuvinte care denumește obiecte (*casă*, *piatră*, *cutie*), substanțe (*apă*, *praf*, *zahăr*), ființe (*student*, *pisică*, *stejar*) sau noțiuni abstracte (*prietenie*, *opinie*, *idee*) iar verbul este definit ca fiind clasa de cuvinte ce exprimă o acțiune (*a alerga*, *a înota*) sau o stare (*a avea*, *a locui*). Alte clase morfologice au o definiție funcțională: pronumele înlocuiește un nume sau un grup nominal (*ieri l-am întâlnit pe Dan*, *cu care* (= *cu Dan*) *am fost coleg*), prepoziția introduce un element subordonat substantivului numit 'atribut' (ca în *casă de piatră*), sau numit 'complement' când este subordonat verbului, adjectivului sau adverbului, (de ex. *mă plimb cu mama*, *se gândește la examene*, *dificil de explicat*), conjuncția este folosită pentru a lega cuvinte sau grupuri de cuvinte (*crede că este bolnav*, *cititorul și scriitorul*), etc.

Analiza în constituenți imediați își propune să dea unităților lingvistice definiții precise, unitare, pe baza formale, în special pe baza capacității de a apărea într-un anumit context, deci pe baza distribuției.

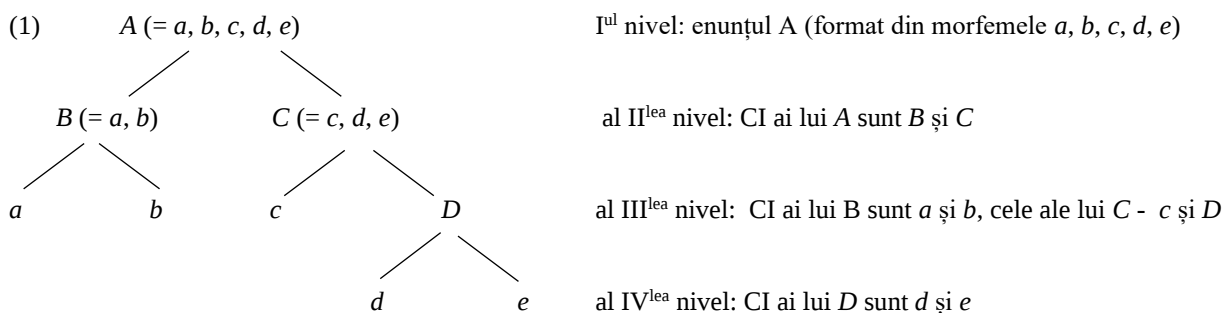
1.1. Principii generale ale analizei în constituenți imediați

Analiza în constituenți imediați, creată de școala distribuționalistă americană, permite să se vadă cu precizie: (i) ierarhia constituenților enunțului și (ii) relațiile de dependență sintactică din interiorul acestor constituenți. Scopul analizei în constituenți imediați este acela de a descompune orice enunț până la nivelul morfemelor, degajând o ierarhie a elementelor ce se îmbină pentru a forma cuvinte, sintagme, propoziții și fraze. Datorită acestei metode, dispare granița dintre morfologie și sintaxă, lingviștii vorbind despre existența unei unice discipline – morfosintaxa (v. MORFOSINTAXĂ).

Se consideră că lingvistul american Kenneth L. Pike (1943) este creatorul metodei analizei în constituenți imediați, într-un studiu publicat în 1943 în revista *Language*, studiu în care autorul a pus în aplicare câteva idei ale lui Leonard Bloomfield (1933). Analiza a fost extinsă și rafinată de o serie de alți lingviști din Statele Unite, dintre care cităm contribuțiile lui Zellig S. Harris (1946), Rulon S. Wells (1947), și Charles F. Hockett (1958).

Termenul de ‘constituent imediat’ (engl. *immediate constituent*) a fost inventat de Bloomfield (1933), care a analizat propoziția *poor John is running away* „bietul Ion fuge (de aici)/ o ia la goană” în șapte elemente (*poor, John, is, run-, -ing, a-, -way*) care sunt ultimii constituenți ai propoziției, adică morfemele ce o constituie (Bloomfield 1933, 161). Dar orice persoana care cunoaște limba engleză își dă seama că propoziția poate fi mai întâi împărțită în două secvențe: *poor John* „bietul Ion” și *is running away* „fuge (de aici)”. La rândul ei, prima secvență poate fi împărțită în *poor* și *John* (sintagmă nominală), și *is running away* (sintagmă verbală), secvență care, la rândul ei, poate fi împărțită în constituenții *is running* (prezentul continuu al verbului *to run* „a fugi, a alerga”) și adverbul *away*; verbul este analizat în *is* (prezentul verbului *to be* „a fi”) și *running* („alergând” participiul prezent, formă constituită din două morfeme *run-* și *-ing*); adverbul *away*, este format tot din două morfeme, *a-* și *-way*.

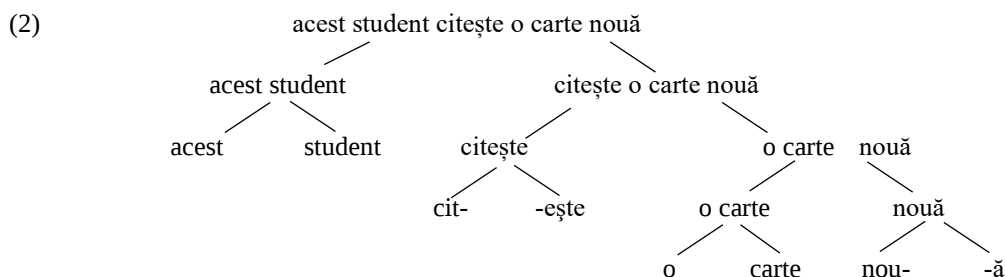
Un lingvist nu poate analiza toate frazele unei limbi, pentru că ele constituie o clasă deschisă și infinită (vezi conceptele chomskiene COMPETENȚĂ și PERFORMANȚĂ). Cercetătorul este, deci, constrâns să supună analizei un număr limitat (mai mare sau mai mic) de fraze, care formează corpusul ales pentru un anumit studiu (vezi CORPUS). Fiecare frază a corpusului este împărțită în două sau trei părți, care, la rândul lor, sunt împărțite succesiv în **constituenți** (lor) **imediați** (CI), până la nivelul morfemelor. Părțile în care propoziția sau fraza poate fi împărțită formează constituenții imediați ai propoziției sau frazei; secvențele rezultate din segmentările următoare sunt constituenții imediați ai nivelului imediat superior:



Dacă neglijăm analiza morfematică a verbului, e ușor de văzut că această structură de constituenți corespunde multor enunțuri cu verb tranzitiv (*acest student citește o carte, orice copil primește un cadou, fiecare spectator cumpără un bilet*, etc.). Deci, după ce este identificată structura sintactică și morfologică a unei anumite propoziții, se constată că aceeași structură caracterizează nu numai enunțul în chestiune, ci o întreagă clasă de enunțuri, care au aceeași structură de constituenți. Secvențele de elemente lingvistice analizate constituie **modele** sau **scheme** (engl. *patterns*) de structuri morfosintactice pentru o anumită limbă. Aceste modele au fost și sunt folosite în predarea unei limbi străine, sub forma exercițiilor structurale.

În comparație cu gramatica tradițională, analiza în constituenți prezintă caracteristicile următoare: (i) se efectuează diviziuni succesive; (ii) pentru fiecare nivel, secvențele care rezultă au același nume - ‘constituenți imediați’; (iii) analiza conduce la dispariția frontierelor dintre sintaxa frazei, sintaxa propoziției și morfologie, deoarece divizările se aplică uniform de la nivelul enunțului, oricât de complex, până la morfemele care îl constituie.

Fie propoziția *acest student citește o carte nouă*. În urma aplicării dicotomiilor succesive (dicotomie = împărțire în două părți) se obține diagrama următoare:



Această structură este în limba română un model pentru o clasă de enunțuri foarte bogată (*acest tânăr a văzut un spectacol interesant, orice elev știe această teoremă fundamentală, fiecare fetiță mănca multe cireșe albe, toți prietenii comentează această întâmplare neobișnuită, etc.*).

Se poate pune întrebarea de ce lingviștii au ales, pentru (2), acest mod de segmentare a propoziției și nu altul, din ce motive a fost aleasa ca primă împărțire dicotomia *acest student* și *citește o carte nouă* când, evident, sunt posibile și alte segmentări.

Pentru alegerea dicotomiei corecte, se aplică două criterii. Conform primului criteriu, divizarea trebuie să conducă la formarea unei **clase distribuționale**, adică la identificarea unui ansamblu de secvențe lingvistice ce pot apărea în același context, deci care sunt echivalente din punct de vedere distribuțional. Este evident că o parte a enunțului poate fi înlocuită cu o parte a unui alt enunț, dacă cele două secvențe fac parte din aceeași clasă de distribuite (v. termenul DISTRIBUȚIE). Dacă examinăm propozițiile *acest student citește o carte nouă* și *Ion aleargă în parc* este evident că segmentele *acest student* și *Ion* pot apărea atât în contextul *__ citește o carte nouă* cât și *__ aleargă în parc*; deci *acest student* și *Ion* sunt echivalente din punct de vedere distribuțional, adică au aceeași distribuție. Asta înseamnă că, dacă propozițiile *acest student citește o carte nouă* și *Ion aleargă în parc* sunt propoziții corecte în limba română, făcând parte din corpusul de analizat, atunci și propozițiile *acest student aleargă în parc* și *Ion citește o carte nouă* sunt de asemenea propoziții corecte ale limbii române.

Numai înlocuirile care conduc la enunțuri reperate (adică prezente în corpus), deci corecte conduc la constituirea unor clase distribuționale. Dacă în propoziție *acest student citește o carte nouă* înlocuim segmentul *citește o carte nouă* cu *repede*, se obține secvența **acest student repede* care nu este o propoziție în limba română. Deci secvențele *citește o carte nouă* și *repede* nu fac parte din aceeași clasă distribuțională. În schimb, *acest student*, *orice tânăr elev*, *Ion* sunt secvențe echivalente din punct de vedere distribuțional, deoarece oricare dintre ele poate apărea în contextul *__citește o carte nouă* iar propozițiile care rezultă sunt reperabile, adică pot fi găsite într-un corpus al limbii române.

Criteriul claselor distribuționale nu este suficient pentru a face segmentări corecte, deoarece întotdeauna sunt posibile mai multe descompuneri. Criteriul claselor distribuționale este respectat nu numai în segmentările din (2), dar și în segmentări ca *acest // student citește o carte nouă*, *acest student citește o // carte nouă*, *acest student citește o carte // nouă* care conduc și ele la enunțuri reperate în limba română, cum ar fi: **orice // student citește o carte nouă**, *acest student citește o // revistă ilustrată*, *acest student citește o carte // interesantă*. Pentru a justifica segmentare *acest student // citește o carte nouă* trebuie să introducem un criteriu suplimentar: este necesar să alegem segmentarea care produce clasa distribuțională cea mai bogată, cu cel mai mare număr de elemente. Deci, numărul de secvențe care pot să apară în contextul *acest student* este mult mai mare decât numărul de secvențe care ar putea apărea în contextul oricărei alte împărțiri. La rândul său, cel de al doilea constituent al propoziției, *__ citește o carte nouă* alcătuiește și el contextul care determină clasa distribuțională cea mai bogată. Astfel se explică faptul că primii constituenți imediați pentru o propoziție ‘standard’ sunt o sintagmă nominală (*acest student*) și o sintagmă verbală (*citește o carte nouă*).

Pentru a clarifica mai bine modul în care cercetătorul alege o anumită dicotomie, Wells (1947) a introdus câteva noțiuni suplimentare. Dacă luăm în considerare două secvențe S_1 și S_2 care au aceeași distribuție, adică pot apărea în aceleași contexte, dar S_2 conține mai multe morfeme decât S_1 , spunem că S_2 este o **expansiune** a lui S_1 ; secvența S_1 se numește **modelul** lui S_2 . Pronumele *ea* constituie modelul unei clase de secvențe constituind sintagme nominale:

- (3) *ea*
 studenta
 studenta inteligentă

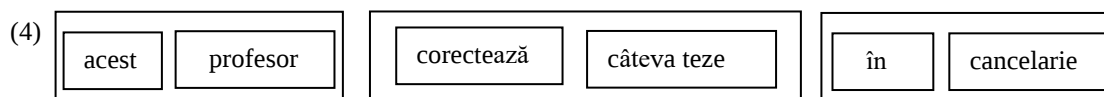
studenta din anul al doilea
studenta pe care toți profesorii o apreciază

clasă alcătuită din elemente cu un număr diferit de morfeme, deci sunt expansiuni posibile ale lui *ea*; toate aceste secvențe au cel puțin o proprietate în comun, aceea de a forma împreună cu secvența *citește o carte nouă* o propoziție corectă în limba română. (Stati, 1966, 32)

1.2. Reprezentarea grafică a ACI

Rezultatele segmentării unei anumite propoziții în constituenți imediați sunt reprezentate grafic ca o succesiune de niveluri ierarhizate. Lingviștii au propus mai multe tipuri de diagrame posibile pentru reprezentarea structurii propoziției: ‘cutii’ (Hockett, 1958, Gleason 1965), paranteze etichetate, arbori.

În reprezentarea prin ‘cutii’ (engl. *Chinese boxes*), fiecare constituent este pus într-un dreptunghi, numit ‘cutie’ – reprezentare inspirată de cutiile chinezești, adică de acele seturi de cutii cu dimensiuni în ordine crescătoare, fiecare cutie mai mică încăpând într-o cutie mai mare. Este, evident, o metaforă a faptului că fiecare constituent este o parte constitutivă a unui constituent mai mare. De exemplu structura enunțului *acest profesor corectează câteva teze în cancelarie* poate fi reprezentată astfel:



Reprezentarea constituenților ‘în cutii’ prezintă câteva dezavantaje: sunt diagrame mai greu de desenat, chiar cu mijloacele moderne, la calculator. În plus, o asemenea reprezentare nu e foarte simplu de citit în cazul unor structuri lingvistice complexe. Nu există nici mijlocul de a indica categoriile cărora le aparțin elementele din interiorul cutiilor.

Reprezentarea prin ‘paranteze etichetate’ (engl. *labeled brackets*) prezintă avantajul de a ocupa mai puțin spațiu, de a fi ușor de scris și de indica prin ‘etichete’ categoria căreia îi aparține fiecare segment al frazei analizate. Parantezele apar în stânga și în dreapta constituentului, arătând în acest fel care îi sunt limitele. La început s-au folosit numai paranteze drepte ‘[’, și ‘]’ (numite în engleză *brackets*), apoi, în interiorul lor, și paranteze rotunde, pentru a reprezenta constituenții subordonați:

(5) [acest profesor] [(corectează) (câteva teze)] [în cancelarie]

Se indică cu o ‘etichetă’ categoria căreia îi aparține fiecare constituent:

(6) [GN acest profesor] [GV (V corectează) (GN câteva tezele)] [GAv în cancelarie]

unde GN = grup nominal, GV = grup verbal, V = verb, GAv = grup adverbial.¹

Dacă dăm toate detaliile, citirea acestei reprezentări e destul de anevoioasă, chiar pentru o propoziție simplă ca (4):

(7) [GN (Det acest) (N profesor)] [GV (V corectează) (GN (Det câteva) (N teze))] [GPrep (Prep în) (N cancelarie)]

De aceea în studiile de lingvistică acest mod de reprezentare a structurii sintactice este folosită pentru construcții relativ simple sau pentru a reprezenta numai principalii constituenți, omițând detaliile:

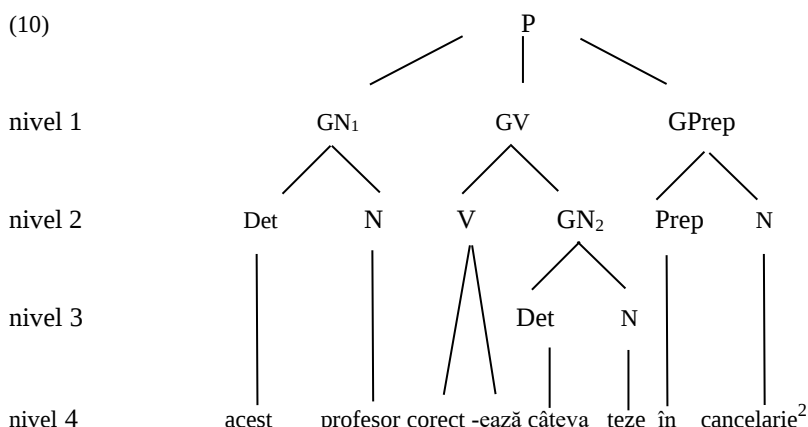
(8) [GN acest profesor] [GV corectează câteva teze)] [GPrep în cancelarie]

¹ Am preluat în linii generale termenii și simbolurile din singura sintaxă transformațională completă a limbii române (Emanuel Vasiliu/ Sanda Golopenția-Eretescu 1969), care a aplicat și adaptat la limba română modelul transformațional ‘clasic’ (Chomsky 1965). Aceiași termeni sunt prezenți și în dicționarele românești de lingvistică (Bidu-Vrâncianu *et al.* 2001). În continuare în text vor mai apărea simbolurile Det pentru ‘determinant’, N pentru ‘nume’ și Prep pentru ‘prepoziție’. De obicei simbolul GAv se folosește pentru complementele circumstanțiale exprimate de un adverb (ca în *a plecat repede*, *a venit devreme*, *a adormit acolo*). Dacă complementul circumstanțial este exprimat de un grup nominal precedat de prepoziție, unii lingviști preferă să folosească simbolul GPrep (grup prepozițional): *a plecat în grabă*, *a venit la ora șase*, *a adormit în fotoliu*, etc.

Se consideră că reprezentarea care dă o imagine a structurii propozițiilor relativ ușor de citit este cea oferită de grafe numite ‘arbori’. În acest tip de diagrame, fiecare constituent, cu excepția simbolului inițial P(ropoziție) este legat de o linie, numită ‘ramură’, de elementul care îi este sintactic subordonat. În mod normal, o propoziție tipică pentru limba română are doi sau trei constituenți, primii doi fiind sintagmele ce formează nucleul propoziției (grupul nominal subiect și grupul verbal ce exprimă predicția), cel de al treilea fiind de obicei un constituent facultativ (un complement ‘circumstanțial’):



Constituenții de (9), de primul nivel, sunt constituenții imediați ai propoziției care sunt analizați în continuare până la ultimii constituenți, morfemele.



Citit de sus în jos, arborele ne permite să ne dăm seama de ierarhia constituenților, începând cu unitatea superioară, P(ropoziție) până la nivelul cuvintelor sau morfemelor. Dacă citim arborele în sens contrar, de jos în sus, urmărim modul în care se grupează secvențele, ca și relațiile care se manifestă în interiorul lor (a se vedea ENDOCENTRIC și EXOCENTRIC). Este evidențiată ierarhia de grupuri sintactice care sunt încorporate, incluse unele în altele, formând unități tot mai mari ce converg spre vârful arborelui, simbolul P.

Un arbore este format dintr-un anumit număr de **noduri**, care sunt legate între ele de **ramuri**. Un nod este un punct de convergență a unei sau mai multor ramuri. În diagrama (10), punctele marcate de simbolurile P, GN₁, GV și GAv/ GPrep sunt noduri. După cum am văzut, aceste simboluri sunt numite **etichete**, adică simboluri ale categoriei sintactice căreia îi aparține secvența deasupra căreia apare. Spunem că anumite unități sunt dominate de un nod dacă ele apar la extremitățile ramurilor ce pornesc din acest nod. Un nod **domină direct** elementele care se află la capătul ramurilor care pornesc de la el și domină indirect elementele care se găsesc la un nivel inferior. În arborele (10), simbolul P domină direct nodurile GN₁, GV, GAV și domină indirect nodurile Det, N, V, GN₂, Prep, N.

Încă de la primele aplicații ale metodei, distribuționaliștii au constatat că anumite secvențe de cuvinte nu se pretează împărțiri binare. Este vorba de așa numitele ‘morfeme discontinue’, care există în toate limbile indo-europene, iar în engleză sunt foarte frecvente. Wells (1947) s-a referit la verbele cu prepoziție din engleză, în construcții de tipul *he took his hat off* „și-a scos pălăria”, unde verbul *take off* „a scoate (un obiect de îmbrăcăminte)” este ‘înterupt’ de prezența complementului direct *his hat* „pălăria lui” (Wells 1947, 200). Gleason (1955) a discutat problema discontinuității în propozițiile interogative din engleză: *did the man come?* „(oare) bărbatul a venit?” trebuie să fie segmentată în *did... come* (forma interogativă a verbului, construită cu

² Pentru a nu complica foarte mult schema, nu am dat toate detaliile morfemelor – de exemplu cuvintele *acest* și *profesor* prezintă un morfem zero (*acest*Ø - *acești*, *profesor*Ø - *profesori*), adjectivul nehotărât *câteva* se poate analiza în adjectivul nehotărât *cât* la masculin plural (*câți*-) și sufixul de compunere -*va* (care apare la pronume, adjective și adverbe: *cineva*, *ceva*, *cândva*, *cumva*, *undeva*, etc.), iar *teze* este o formă de plural a substantivului (*tez-ă* - *tez-e*). Substantivul *cancelarie* prezintă de asemenea două morfeme (*cancelari-e* - *cancelari-i*). În limba română, identificarea morfemelor trebuie să înfrunte în plus problema, de loc ușoară, a alternanțelor la nivelul lexemului (*profesor* - *profesoară*, *câtiva* - *câteva*, etc.).

perfectul auxiliarului *to do - did* și subiectul, grupul nominal *the man* (Gleason 1955, 142). La nivelul simplei ACI, problema a fost semnalată fără să fie rezolvată.

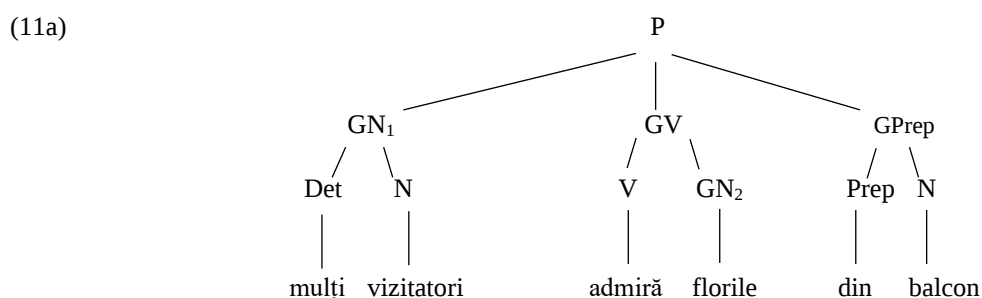
Probleme de discontinuitate apar și în aplicarea analizei în CI la limba română: de exemplu așa-numitul ‘gen personal’ se manifestă, la acuzativ, prin dublare pronominală și prezența ‘prepoziției’ *pe* (*Îl văd pe Ion*).

1.3. Ambiguitatea sintactică

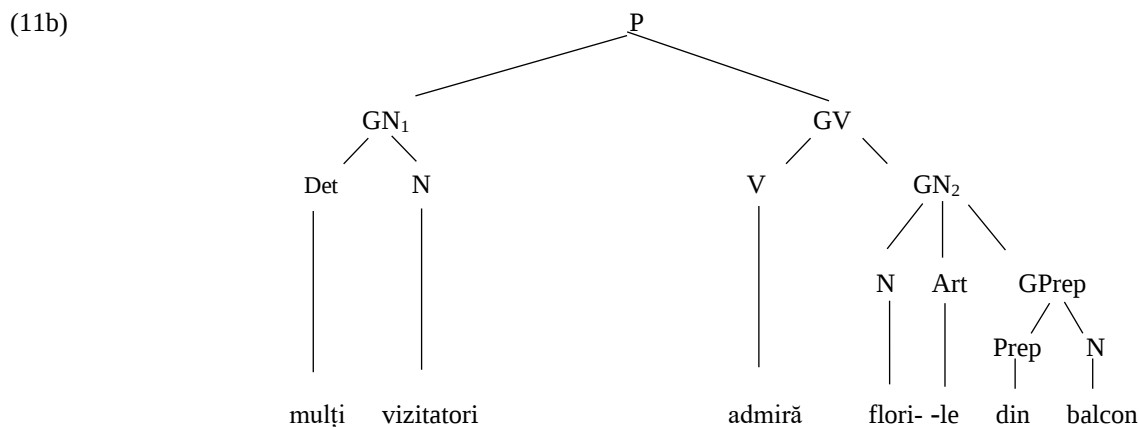
Chomsky (1965, 21-22) a indicat modul în care reprezentarea structurii frazei permite să se explicitizeze anumite ambiguități a unor construcții, deci existența unei omonimii sintactice. De exemplu o propoziție ca

(11) Mulți vizitatori admiră florile din balcon.

este ambiguă, poate avea două interpretări: vizitatorii se găsesc pe balcon și de acolo privesc florile care se găsesc în altă parte, de exemplu în grădina de sub balcon (în acest caz, grupul prepozițional *din balcon* este complement circumstanțial de loc, sintagmă dominată direct de simbolul P):



Într-o altă interpretare, propoziția spune că florile se găsesc pe balcon și vizitatorii se găsesc în altă parte, de exemplu pe strada de sub balcon (în acest caz grupul prepozițional *din balcon* este atribut prepozițional, dominat de GN₂):



După Chomsky (1965), gramatica de constituenți are meritul de a înregistra aceste fenomene sintactice, dar simpla identificare a construcțiilor sintactic ambigue nu este suficientă, este necesar și să fie explicată. Este un argument folosit de Chomsky în favoarea necesității de a introduce în gramatica generativă a regulilor de transformare.

2. Gramatica generativă de constituenți

Noam Chomsky (1956, 1957, 26-33) a pornit de la analiza în constituenți imediați pentru a formula prima sa variantă de gramatică generativă, numită ‘de constituenți’.³ După cum am văzut, analiza (atât cea

³ Prima gramatică generativă în absolut a fost așa-numita ‘gramatică cu număr finit de stări’ sau ‘automat Turing’, de la numele creatorului ei, matematicianul britanic Alan Turing (1912-1954), logician, filozof, criptograf, unul dintre părinții informaticii moderne.

tradițională cât și cea structurală sub forma ACI) pornește de examinarea unui anumit număr de enunțuri, pentru a identifica diferitele categorii și stabilirea modul în care acestea se comportă (forme morfologice și funcțiile lor sintactice în gramatica tradițională, clasele distribuționale, modele și expansiuni în ACI), comportament descris de așa-numitele reguli. Acest parcurs ilustrează un demers analitic, ce pornește de la enunțurile limbii: în urma studiului, cercetătorul stabilește regulile morfologice și sintactice care descriu comportamentul lingvistic al unităților identificate (clase morfologice cu anumite funcții sintactice sau constituenți imediați). O gramatică generativă parcurge traseul invers: pornește de la un număr de reguli prin aplicarea cărora se ajunge la producerea, adică la generarea unui număr de enunțuri.

În gramatica generativă de constituenți, principala regulă este o regulă numită ‘de rescriere’ având forma:

$$(12) \quad A \rightarrow B + C \quad (\text{instrucțiune: rescrie/ înlocuiește simbolul A prin succesiunea de simboluri B și C})$$

Regulile de rescriere, care îndeplinesc funcția de a genera enunțuri, se prezintă într-ună formă ce dovedește că Chomsky (1957) a ‘răsturnat’ rezultatele procesului analitic, reprezentat de ACI, pentru a crea o gramatică sintetică, generativă:

$$(13) \quad \begin{aligned} P &\rightarrow GN + GV \\ GN &\rightarrow N + \text{Art}_h (= \text{articol hotărât}) \\ V &\rightarrow V + GN \\ N &\rightarrow \textit{student, profesor, carte, revistă} \\ \text{Art}_h &\rightarrow \textit{-ul, -a} \\ V &\rightarrow \textit{citi, răsfoi, studia} \end{aligned}$$

Prin aplicarea acestor reguli, fragmentul de gramatică de constituenți de la (13) poate genera propoziții ca *studentul citește carte, profesorul răsfoiește revista, studentul studiază cartea*, etc. Procesul de generare se face prin aplicarea regulilor de rescriere, de fiecare dată fiind înlocuit un singur element. Acest tip de succesiune a operațiilor care conduce la generarea unui enunț se numește **derivație**:

$$(14) \quad \begin{aligned} &P \\ &GN + GV \\ &N + \text{Art}_h + GV \\ &N + \text{Art}_h + V + GN \\ &N + \text{Art}_h + V + N + \text{Art}_h \\ &\textit{profesor} + \text{Art}_h + V + N + \text{Art}_h \\ &\textit{profesor} + \textit{-ul} + V + N + \text{Art}_h \\ &\textit{profesor} + \textit{-ul} + \textit{studiază} + N + \text{Art}_h \\ &\textit{profesor} + \textit{-ul} + \textit{studiază} + \textit{carte} + \text{Art}_h \\ &\textit{profesor} + \textit{-ul} + \textit{studiază} + \textit{carte} + \textit{-a}^4 \end{aligned}$$

Chomsky (1965) critică gramatica generativă de constituenți, care reprezintă un mare progres față de gramatica cu număr finit de stări dar nu are o putere explicativă suficientă, pentru a fi un model al competenței lingvistice a vorbitorilor.

Analiza în constituenți imediați este în măsură să explice unele cazuri ambiguitate sintactică, adică faptul că un anumit enunț poate avea două descrieri structurale, fiecare corespunzând unei anumite semnificații, după cum am văzut pentru propoziția (11) care poate fi descrisă de două structuri diferite de constituenți (11a) și (11b). Există însă situații de ambiguitate pe care funcțiile sintactice și prezentarea structurii lor de constituenți nu reușesc să le explice. De exemplu, o sintagmă ca *primirea prietenilor* este alcătuită de un substantiv regent și un atribut substantival genitival. Sintagma este ambiguă, atributul genitival oscilând între o interpretare ‘subiectivă’ (*prietenii primesc (pe cineva)*) și o interpretare ‘obiectivă’ (*(cineva) îi primește pe prieteni*) (Vasiliu/ Golopenția 1969, 24). Ambiguitățile sintactice sunt relativ frecvente: *critica lui Bloomfield* (care poate deriva de la *(cineva) îl critică pe Bloomfield* sau de la *Bloomfield critică pe cineva/ ceva*), *Ion vede o fată trecând pe stradă* (ambiguitate în privința agentului verbului *a trece* care poate fi *Ion* (*Ion a văzut o fată în timp ce (Ion) trecea pe stradă*) sau *o fată* (*Ion a văzut o fată care trecea pe stradă*), etc.

ACI nu poate explica nici cazurile de parafrază sau ‘sinonimie’ sintactică, adică structuri de constituenți diferite care transmit aceeași informație. Acest tip de analiză nu poate lămuri nici legătura evidentă

⁴ Regulile care asigură forma articulată corectă a substantivelor sau acordul verbului cu grupul nominal subiect sunt asigurate de un tip special de reguli, numite ‘morfofonemice’, pe care nu le precizăm pentru a simplifica prezentarea.

dintre *elevul citește carte* și *cartea este citită de elev* (corespondent pasiv), *invitații vin* și *venirea invitaților*, etc.

Chomsky (1965) a semnalat aceste situații pe care gramaticile generative de constituenți nu le explică în mod satisfăcător ca argument pentru necesitatea îmbogățirea sintaxei generative cu reguli de transformare care convertesc o structură de adâncime (produsă de o gramatică generativă de constituenți, ca cea din (13) - (14)) într-o structură de suprafață.

Cazurile de omonimie (sau ambiguitate) sintactică se explică prin existența a două procese diferite de generare, pornind de la structuri de adâncime diferite: transformarea de nominalizare poate fi aplicată verbului *a primi* atât într-o structură de adâncime apropiată de (*cineva*) *îi primește pe prieteni* cât și *prietenii primesc (pe cineva)*. Avem structuri de adâncime diferite și pentru sintagmele *critica lui Bloomfield* ((*cineva*) *l-a critică pe Bloomfield* sau *Bloomfield a criticat (pe cineva)*), *trecând pe stradă* (*Ion trece pe stradă* sau *fata trece pe stradă*), etc. Tot așa se explică ambiguitatea enunțului (11), care poate fi generat pornind de la două structuri de adâncime diferite, una apropiată de (11a), alta de (11b).

Cazurile de ‘sinonimie’ sintactică se explică invers, prin existența unei singure structuri de adâncime ulterior supusă la transformări diferite. De exemplu, în structura de adâncime a sintaxei propusă în Chomsky (1965) nu există propoziții la forma pasivă; pasivul este obținut prin aplicarea transformării pasive. Deci *elevul citește carte* și *cartea este citită de elev* au aceeași structură de adâncime, cea de doua formă rezultând în urma aplicării transformării pasive asupra primei forme, active. În aceeași sintaxă există transformarea de nominalizare, ceea ce semnifică că substantivele nominalizate nu apar în structura de adâncime. Deci *venirea invitaților* este o sintagmă din structura de suprafață care rezultă din aplicarea transformării de nominalizare la o structură de adâncime de tip *invitații vin*, ceea ce explică legătura semantică cele două enunțuri.

3. Importanța analizei în constituenți imediați

Analizei în constituenți imediați, numită și ‘analiză distribuțională’, a avut o importanță excepțională în dezvoltarea lingvisticii în a doua jumătate a secolului trecut, în special în evoluția sintaxei. Vom arăta trei domenii în care această analiză a avut un rol important.

Modul în care noi toți am avut contact direct cu rezultatele acestui tip de studiu derivă din faptul că acest model de analiză a fost mult aplicat în didactica limbilor străine, fiind la baza așa-numitelor ‘exerciții structurale’ pe care le regăsim în aproape toate manualele de limbi străine ulterioare anului 1960.

Comentarea, critica și renovarea analizei în constituenți imediați se găsește în primele două cărți ale lui Chomsky (1957, 1965). Pornind de la acest tip de analiză morfosintactică formală, Chomsky a reușit acea ‘răsturnare copernicană’ din lingvistică, adică trecerea de la o gramatică analitică la o gramatică sintetică, generativă. Ulterior, gramatica generativă a adoptat un nou mod de reprezentare a ierarhiei structurilor sintactice, mai general, universal valabil - așa numita reprezentare ‘X-bară’ (Chomsky 1982).

Dezvoltarea gramaticilor generative din anii ’70 ai secolului trecut este contemporană cu dezvoltarea explozivă a informaticii și a computerelor, atât la nivelul dispozitivelor tehnice, al ordinatorilor, cât și al programelor de operare. Analiza în constituenți imediați a fost folosită de informaticieni și de specialiști în domeniul interdisciplinar al informaticii lingvistice, așa numita ‘lingvistică computațională’ (engl. *computational linguistics*). Analiza în constituenți imediați se găsește la baza a numeroase aplicații de corectare automată a diferitelor tipuri de erori, a diferitelor analize cantitative și de corpus. Desigur că cea mai spectaculoasă aplicație este cea a traducerii automate care, la nivelul sintaxei, a folosit gramatici de constituenți imediați (Parker-Rhodes, 1966).

Bibliografie

- Bidu-Vrînceanu, Angela, Cristina Călărașu, Liliana Ionescu-Ruxăndoiu, Mihaela Mancaș, Gabriela Pană Dindelegan, (2001), *Dicționar de științe ale limbii*, București, Nemira.
- Blevins, James/ Ivan Sag (2013), ‘Phrase Structure Grammar’, in Marcel den Dikken (ed.) *The Cambridge Handbook of Generative Syntax*, 202-225.
- Bloomfield, Leonard (1933), *Language*, New York, Henry Holt.
- Chomsky, Noam (1956), ‘Three models for the description of language’, *Institute of Radio Engineers Transactions on Information Theory*, II-2, 113-124. Reprinted in Luce et al. (1965), *Readings in Mathematical Psychology* 2, Wiley and Sons, New York, 105 - 124.
- Chomsky, Noam (1957), *Syntactic Structures*, The Hague, Mouton
- Chomsky, Noam (1965), *Aspects of the Theory of Syntax*, Cambridge, The MIT Press.
- Chomsky, Noam (1982), *Lectures on Government and Binding*, Dordrecht, Fortis.
- Harris, Zellig S. (1946), ‘From Morpheme to Utterance’, in *Language* 22, 161-183.

- Hockett, Charles F. (1958), 'A course in modern linguistics', in *Language Learning*, 8(3-4), 73-75.
- Parker-Rhodes, Arthur F. (1966) 'Syntax' in *Automatic Translation of Languages*, Papers presented at NATO Summer School held in Venice, July 1962, Amsterdam, Elsevier Ltd., 167-203; disponibil la <<https://aclanthology.org/1962.earlymt-1.7.pdf>>.
- Pike, Kenneth L. (1943), 'Taxemes and immediate constituents', in *Language*, 65-82.
- Stati, Sorin (1966), 'Lingvistica structurală', în Solomon Marcus/ Edmond Nicolau/ Sorin Stati (eds.) *Introducere in lingvistica matematică*, București, Editura științifică, 7-41.
- Vasiliu, Emanuel/ Sanda Golopenția-Eretescu (1969), *Sintaxa transformățională a limbii române*, București, Editura Academiei RSR.
- Wells, Rulon S. (1947), 'Immediate constituents', in *Language*, 23(2), 81-117.