

কম্পিউটাৰ ভাষাবিজ্ঞান

কম্পিউটাৰ ভাষাবিজ্ঞান হ'ল ভাষাবিজ্ঞান আৰু কম্পিউটাৰ বিজ্ঞানৰ সংযুক্তিত গঢ়ি উঠা আন্তৰ্জালিত এটা বিষয় যি মানুহৰ স্বাভাৱিক ভাষিক ক্ষমতাক কম্পিউটাৰত প্ৰয়োগৰ আলোচনা কৰে। গ্ৰিছমেনে কম্পিউটাৰ ভাষাবিজ্ঞানৰ সংজ্ঞা দিছে এনেদৰে— কম্পিউটাৰ ভাষাবিজ্ঞান হ'ল ভাষা বুজা আৰু উৎপাদন বাবে কম্পিউটাৰ পদ্ধতিৰ অধ্যয়ন। (Grishman (1986) defines Computational linguistics as the study of computer systems for understanding and generating natural language.)

এই সংজ্ঞাৰ পৰা আমি বুজিব পাৰোঁ যে কম্পিউটাৰ ভাষাবিজ্ঞানে প্ৰধানকৈ দুটা কাম কৰে—

(১) মানুহৰ স্বাভাৱিক ভাষাসমূহক বুজিবলৈ প্ৰযত্ন কৰে, আৰু

(২) মানুহৰ দৰে স্বাভাৱিক ভাষা উৎপাদনৰ প্ৰচেষ্টা কৰে।

কিন্তু এই প্ৰক্ৰিয়া যথেষ্ট জটিল আৰু কষ্টসাধ্য। কিয়নো মানুহে ব্যৱহাৰ কৰা স্বাভাৱিক ভাষাসমূহ যথেষ্ট জটিল, যদিও এইবোৰ গঠনৰ নিৰ্দিষ্ট কেতবোৰ নিয়ম আছে। এই নিয়মসমূহক আমি ব্যাকৰণ বুলি কওঁ। ভাষাৰ এনে নিয়ম যথেষ্ট সীমিত যদিও এই সীমিত নিয়মেৰে আমি অসংখ্য বাক্য সৃষ্টি কৰোঁ।

আনহাতে, কম্পিউটাৰে যিহেতু এটা মেছিনহে আৰু ই মাত্ৰ কেৱল কেতবোৰ গাণিতিক সূত্ৰৰ দ্বাৰাহে পৰিচালিত হয়। ইয়াৰ অৰ্থ এইটো যে ব্যাকৰণৰ সূত্ৰসমূহ গাণিতিক সূত্ৰলৈ ৰূপান্তৰ কৰা। ভাষাৰ স্বাভাৱিক প্ৰক্ৰিয়াকৰণৰ () জৰিয়তে কম্পিউটাৰৰ ভাষালৈ অৰ্থাৎ গাণিতিক সূত্ৰলৈ ৰূপান্তৰৰ জৰিয়তে কম্পিউটাৰ ভাষাবিজ্ঞানৰ কাৰ্য পৰিচালিত হয়। অৰ্থাৎ, পদ্ধতিগত ৰূপত ভাষা প্ৰক্ৰিয়াক বুজি, গাণিতিক পদ্ধতিৰে স্বাভাৱিক ভাষা উৎপাদন আৰু ব্যাখ্যা কৰাৰ সক্ষমতা কম্পিউটাৰক প্ৰদান কৰিব পাৰোঁ। ইয়ে ভাষিক কাৰ্য (যেনে—যন্ত্ৰ অনুবাদ), পাঠৰ তথ্য প্ৰক্ৰিয়া (পুথি, জাৰ্ণাল, বাতৰিকাকত), পৰিচালনাৰ বাবে আৰু কম্পিউটাৰত সংৰক্ষিত তথ্যসমূহক আহৰিত (access) কৰাত সহজসাধ্য কৰি তুলিব পাৰে। ভাষা ব্যৱহাৰৰ এনে সুসংহত সক্ষমতাৰ জৰিয়তে এক আমূল পৰিৱৰ্তন সাধন কৰিব পৰা যায়।

কম্পিউটাৰ ভাষাবিজ্ঞানৰ গৱেষণা বা অনুসন্ধানৰ এক ব্যাপক ক্ষেত্ৰ আছে যদিও প্ৰাথমিক প্ৰেৰণা সদায় প্ৰাকৃতিক বা স্বাভাৱিক ভাষাক বিশিষ্ট ব্যৱহাৰিক প্ৰণালীৰে ব্যৱহাৰ কৰিব পদ্ধতিৰ ওপৰতে দিয়া হৈ আহিছে। কম্পিউটাৰ ভাষাবিজ্ঞানত তিনিটা ক্ষেত্ৰক প্ৰধানকৈ গুৰুত্ব দিয়া হৈছে—

যন্ত্ৰানুবাদ (Machine translation)

তথ্য আহৰণ (Information retrieval)

মানুহ-মেছিনৰ পাৰস্পৰিক আদান-প্ৰদান (Man-machine interfaces)

ইয়াৰ বাবে আমাক স্বাভাৱিকভাৱেই কম্পিউটাৰ ভাষাবিজ্ঞানত প্ৰয়োজন হ'ব তাত্ত্বিক ব্যাকৰণবিদসকলে দাঙি ধৰা ব্যাকৰণৰ পৰীক্ষণ। ৰূপান্তৰস্ফোট ব্যাকৰণ (Transfomal Generative Grammar) ব্যৱহাৰৰ জৰিয়তে কম্পিউটাৰ ভাষাবিজ্ঞানত ভাষা একোটাৰ জটিল নিয়মসমূহক পৰীক্ষা কৰাৰ উপযুক্ত মাধ্যম বুলি পৰিগণিত হ'ব পাৰে। Friedman's Transformational Grammar Tester (Friedman J 971) এনে এক পদ্ধতি যি ৰূপান্তৰস্ফোট ব্যাকৰণৰ গঠনপদ্ধতিৰ দৰে বাক্য গঠন কৰে। সংজ্ঞানাত্মক বিজ্ঞান (cognitive sciences) আৰু কৃত্ৰিম বুদ্ধিমত্তা (artificial intelligence) ৰ লগত আংশিক পৰিমাণে সম্পৰ্কিত কম্পিউটাৰ বিজ্ঞানৰ এনে এক শাখা যি মানৱ অনুভূতিক কম্পিউটাৰৰ আৰ্হিৰ সহায়ত প্ৰকাশ কৰে। কম্পিউটাৰ ভাষাবিজ্ঞানৰ নিজস্ব প্ৰায়োগিক আৰু তাত্ত্বিক উপাদান আছে।

তাত্ত্বিক কম্পিউটাৰ ভাষাবিজ্ঞানে ভাষাবিজ্ঞানৰ সূত্ৰ আৰু সংজ্ঞানাত্মক বিজ্ঞানৰ বিষয়সমূহক আধাৰ হিচাপে গ্ৰহণ কৰে। মানুহ এজনে ভাষা কেনেকৈ উৎপাদন কৰে আৰু বুজি পায় সেয়া বুজিবৰ কাৰণে প্ৰয়োজন হোৱা আনুষ্ঠানিক সূত্ৰসমূহৰ লগত ই জড়িত। এই সূত্ৰসমূহ আজি এনে এক জটিল মাত্ৰা পাইছেগৈ যে সেইবোৰ বিশ্লেষণ কেৱল কম্পিউটাৰ প্ৰবন্ধনৰ জৰিয়তেহে কৰিব পৰা যায়। কম্পিউটাৰ ভাষাবিজ্ঞানীসকলে মানুহৰ ভাষাসমূহৰ বিভিন্ন দিশসমূহৰ ভিত্তিত কেতবোৰ আনুষ্ঠানিক আৰ্হি (model)-ৰ বিকাশ কৰিছে আৰু কম্পিউটাৰ কাৰ্যক্ৰম (programs) হিচাপে সিৰোৰক ব্যৱহাৰ কৰিছে। এই কাৰ্যক্ৰমবোৰৰ মূল্যায়ন আৰু তাত্ত্বিক ভিত্তিত অনাগত দিনত অধিক উন্নত তত্ত্বৰ বিকাশ ঘটিব। ভাষাবিজ্ঞানৰ সূত্ৰসমূহৰ উপৰি সংজ্ঞানাত্মক মনোবিজ্ঞানৰ (cognitive psychology)ৰ পৰা পোৱা ফলাফলসমূহে ভাষিক দক্ষতা অনুকৰণত গুৰুত্বপূৰ্ণ ভূমিকা পালন কৰে। ই মূলতঃ মনোভাষাবিজ্ঞানৰ ক্ষেত্ৰ যি মানৱ ভাষাৰ গঠনৰ সংজ্ঞানাত্মক প্ৰক্ৰিয়াসমূহ পৰীক্ষা কৰে। মনোভাষাবিজ্ঞানৰ গৱেষণাৰ কাৰণে কম্পিউটাৰীয় আৰ্হিয়ে এক নতুন উপশাখা উদ্ভৱ কৰে— কম্পিউটাৰ মনোভাষাবিজ্ঞান (Computational psycholinguistics)। কম্পিউটাৰ ভাষাবিজ্ঞানে সামৰি লোৱা বিষয়বোৰ চিত্ৰৰ সহায়ত ব্যাখ্যা কৰিলে এনেধৰণৰ হ'ব—



প্ৰায়োগিক কম্পিউটাৰ ভাষাবিজ্ঞান বা প্ৰায়োগিক সংগণনীয় ভাষাবিজ্ঞানে ভাষা ব্যৱহাৰৰ ব্যৱহাৰিক ফলাফলৰ ওপৰত গুৰুত্ব প্ৰদান কৰে। পদ্ধতি (method), প্ৰযুক্তিগত কৌশল (techniques), আহিলা (tools) আৰু প্ৰয়োগ (applications) আদি ভাষিক কৌশল উদ্ভাৱনৰ (language engineering) বা ভাষিক প্ৰযুক্তিবিদ্যা (language technology) শৃংখলাৰ অধীনস্থ ক্ষেত্ৰৰ লগত জড়িত। বৰ্তমান যদিও সংগণনীয় ভাষাবিজ্ঞান পদ্ধতিসমূহ মানুহৰ দৰে দক্ষতা অৰ্জন ক্ষমতাৰ পৰা অনেক দূৰত তথাপি ইয়াৰ সন্মুখত আছে প্ৰায়োগিক দিশত অনেক সম্ভৱনা। লক্ষ্য হ'ল মানুহৰ দৰে জ্ঞানসম্পন্ন ছফ্টৱেৰৰ নিৰ্মাণ কৰা যিবোৰে মানুহৰ ভৱিষ্যৎ জীৱন সলনি কৰি দিব পাৰে। তাৎক্ষণিক ৰূপত মানুহ আৰু কম্পিউটাৰৰ মাজৰ পাৰস্পৰিক ভাব বিনিময়ৰ মাধ্যমৰ বিকাশৰ প্ৰয়োজন, কিন্তু মূল বাধা হৈ থিয় দিছে মানুহ আৰু কম্পিউটাৰৰ মাজৰ সংযোগহীনতাৰ সমস্যা। বৰ্তমান সময়ত মানুহৰ ভাষা কম্পিউটাৰে বুজি নাপায়, বা কম্পিউটাৰৰ ভাষাও যথেষ্ট কষ্টসাধ্য আৰু মানুহৰ চিন্তাধাৰাৰ কাঠামোৰ লগত সম্পৰ্কিত নহয়।

কম্পিউটাৰ ভাষাবিজ্ঞানৰ উদ্দেশ্য হ'ল মানৱীয় ভাষাক কম্পিউটাৰত প্ৰয়োগ (applications)ৰ বিকাশ ঘটোৱা, যেনে— ব্যাকৰণ সংশোধন ছফ্টৱেৰৰ বিকাশ, যথায়থ শব্দৰ প্ৰয়োগ (word sense disambiguation), শব্দকোষ আৰু পাঠ নিৰ্মাণ, আৰু এটা ভাষাৰ পৰা আন এটা ভাষালৈ স্বতস্ফুৰ্তঃ যন্ত্ৰানুবাদ আদি।

কম্পিউটাৰ বিজ্ঞানৰ আৰম্ভণি

কম্পিউটাৰ ভাষাবিজ্ঞানৰ আৰম্ভণি হয় আমেৰিকাত ১৯৫০ চনৰ পৰা। ৰাছিয়াৰ বিজ্ঞান গৱেষণা পত্ৰিকা কম্পিউটাৰৰ দ্বাৰা স্বতস্ফুৰ্তঃ যন্ত্ৰানুবাদৰ প্ৰচেষ্টাত কাৰ্যতঃ এই কম্পিউটাৰ ভাষাবিজ্ঞানৰ আৰম্ভণি হয়। অধ্যয়ন আৰু গৱেষণাৰ দ্বাৰা প্ৰধানকৈ ভাষাৰ সূত্ৰসমূহক প্ৰণালীবদ্ধ ৰূপত উপস্থাপন কৰি algorithms আৰু software-ৰ বিকাশ সাধনৰ উদ্দেশ্যেই এই নতুন ক্ষেত্ৰ উৎসৰ্গিত।

কম্পিউটাৰৰ জৰিয়তে স্বয়ংক্ৰিয় বা যন্ত্ৰানুবাদৰ প্ৰচেষ্টা আৰম্ভ হোৱাৰে পৰাই এই ক্ষেত্ৰ ভাষাবিজ্ঞানৰ লগত সম্পৰ্কিত হয়। কোনো এটা পাঠৰ অনুবাদ কৰিবলৈ হ'লে, দুয়োটা ভাষাৰ ব্যাকৰণ বা গঠন জানিব লাগিব। অৰ্থাৎ অনুবাদকজনে উৎস ভাষা আৰু লক্ষ্য ভাষাৰ ৰূপতত্ত্ব, বাক্যতত্ত্ব, অৰ্থ, প্ৰসঙ্গাৰ্থ আদি নিৰহ নিপানিকৈ জানিব লাগিব।

কম্পিউটাৰ ভাষাবিজ্ঞানৰ প্ৰস্তাৱিত ক্ষেত্ৰঃ—

কম্পিউটাৰ ভাষাবিজ্ঞান বৰ্তমান চালুকীয়া অৱস্থাতে আছে। এই বিষয়ত অৱশ্যে অহৰহ অধ্যয়ন আৰু গৱেষণা চলিছে, যাতে কম্পিউটাৰ ভাষাবিজ্ঞানৰ সহায়েৰে মানুহৰ জীৱন ধাৰণা পদ্ধতিক অতি সহজ কৰিব পৰা যায়। মানুহৰ স্বাভাৱিক ভাষা যথেষ্ট জটিল। আকৌ ভাষা বিশেষে গঠন পদ্ধতিও বেলেগ বেলেগ। সেয়েহে ভাষাবিজ্ঞানী আৰু কম্পিউটাৰ বিজ্ঞানীসকলে অহৰহ চেষ্টা চলাইছে যাতে ভাষাৰ গঠনক প্ৰণালীবদ্ধ ৰূপত উপস্থাপন কৰিবলৈ। সেয়ে প্ৰাথমিকভাৱে গুৰুত্ব দিয়া বিষয় হ'ল—

১. তত্ত্ব-আধাৰিত পদ্ধতি (Rule-Based Systems) —

- (ক) ভাষিক জ্ঞানক পদ্ধতিগত ৰূপত প্ৰকাশ,
- (খ) ব্যাকৰণৰ নিয়মসমূহক সমষ্টি ৰূপত উপস্থাপন,

কম্পিউটাৰ ভাষাবিজ্ঞানৰ দ্বিতীয় গুৰুত্বপূৰ্ণ দিশটো হ'ল—

ডাটা-চালন পদ্ধতি (Data-Driven Systems)

- (ক) ভাষাৰ অন্তৰ্নিহিত জ্ঞান (Implicit encoding of linguistic knowledge)
- (খ) পৰিসংখ্যান পদ্ধতি বা যন্ত্ৰ-শিকন পদ্ধতিৰ প্ৰয়োগ

কম্পিউটাৰ ভাষাবিজ্ঞানৰ ক্ষেত্ৰ

কম্পিউটাৰ ভাষাবিজ্ঞানৰ ক্ষেত্ৰসমূহ প্ৰধানকৈ—

- (ক) মেছিন অনুবাদ বা যন্ত্ৰানুবাদ (Machine Translation)
- (খ) বাক-স্বীকৃতকৰণ (speech recognition)
- (গ) মানুহ-যন্ত্ৰৰ পাৰস্পৰিক প্ৰদান প্ৰদান (man-machine interfaces)
- (ঘ) ব্যাকৰণ শুদ্ধিকৰণ (Grammar Checker)
- (ঙ) বানান শুদ্ধিকৰণ (Spelling Checker)

অদূৰ ভৱিষ্যতে কম্পিউটাৰ ভাষাবিজ্ঞানৰ জৰিয়তে অনেক পৰিৱৰ্তন আশা কৰা হৈছে। কম্পিউটাৰ ভাষাবিজ্ঞানৰ বিকাশৰ জৰিয়তে কম্পিউটাৰক অধিক কাৰ্যক্ষম কৰি তোলাৰ পৰিকল্পনা কৰা হৈছে।

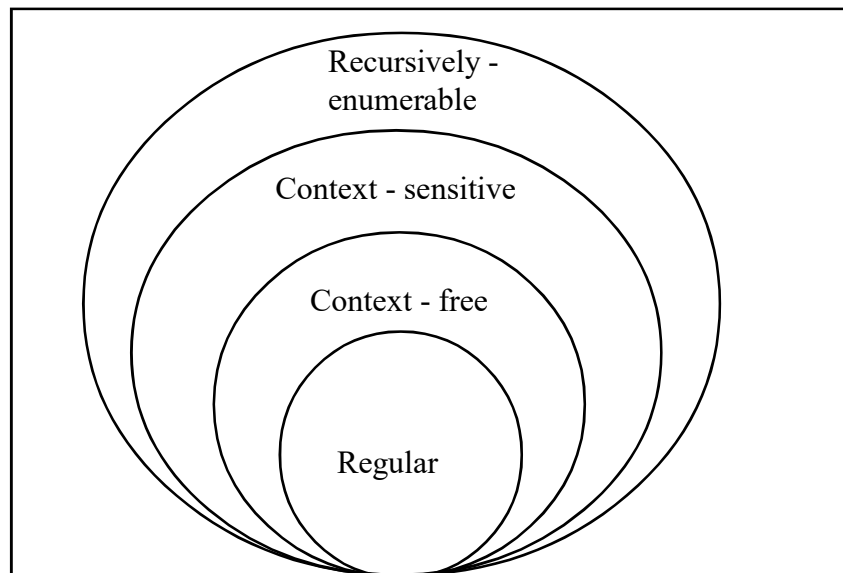
a set of nonterminal symbols, which are placeholders for patterns of terminal symbols that can be generated by the nonterminal symbols..

প্রসঙ্গ মুক্ত ব্যাকরণ (Context-Free Grammars) \Rightarrow

চমস্কি মত অনুসৰি ব্যাকৰণ চাৰি প্ৰকাৰৰ— Type 0, Type 1, Type 2, and Type 3 । তলত এই ব্যাকৰণসমূহে কি ধৰণৰ ব্যাকৰণ আৰু ভাষা গ্ৰহণ কৰে আৰু লগতে মেছিনসমূহে ব্যৱহাৰ কৰা মেছিনৰ উল্লেখ কৰা হ'ল—

Grammar Type	Grammar Accepted	Language Accepted	Automaton
Type 0	Unrestricted grammar	Recursively enumerable language	Turing Machine
Type 1	Context-sensitive grammar	Context-sensitive language	Linear-bounded automaton
Type 2	Context-free grammar	Context-free language	Pushdown automaton
Type 3	Regular grammar	Regular language	Finite state automaton

তলৰ চিত্ৰত প্ৰত্যেকটো ধৰণৰ ব্যাখ্যা আগবঢ়োৱা হৈছে—



Type-3 ব্যাকৰণে regular languages উৎপন্ন কৰে। Type-3 ব্যাকৰণত বাওঁফালে এটা non-terminal থাকিব লাগিব আৰু সোঁফালে এটা terminal বা non-terminal এটাক terminal-এ এটাই অনুসৰণ কৰে।

উৎপাদনটো নিশ্চিতকৈ এই ৰূপত থাকিব লাগিব— $X \rightarrow a$ or $X \rightarrow aY$

য'ত $X, Y \in N$ (Non terminal)

আৰু $a \in T$ (Terminal)

$S \rightarrow \epsilon$ নিয়মটো অনুমোদিত যদিহে S সোঁফালে কোনো নিয়মতে দেখা নাযায়।

Example

```
X → ε
X → a | aY
Y → b
```

Type-2 ব্যাকৰণৰ (প্ৰসঙ্গমুক্ত) দ্বাৰা উৎপন্ন ভাষাসমূহক কোৱা হয় প্ৰসঙ্গমুক্ত ভাষা (Context-free languages)। নিয়মসমূহৰ মাজত সীমাবদ্ধতা আছে যাৰ ফলত সেইবোৰ $A \rightarrow \gamma$ ৰূপত উৎপন্ন হ'ব।

য'ত $A \in N$ (Non terminal)

আৰু $\gamma \in (T \cup N)^*$ (String of terminals and non-terminals).

ব্যাকৰণৰ সীমাবদ্ধতা থকা বাবে বাওঁফালে এটা মাত্ৰ nonterminal চিহ্ন থাকিব আৰু কোনো প্ৰসঙ্গ নাথাকে। এইটোৱে মূলতঃ সূচায় যে যদি এটা অন্তৰ্ৱৰ্তী শব্দত nonterminal চিহ্ন থাকে, তেতিয়া তাত কোনো এটা নিয়ম প্ৰয়োগ কৰি বাওঁফালে থকা nonterminal চিহ্নক সোঁফালে প্ৰতিস্থাপন কৰিব পাৰি। nonterminal চিহ্নৰ পৰিৱেশক উপেক্ষা কৰি। অধিকাংশ প্ৰ'গ্ৰেমিং ভাষা প্ৰসঙ্গমুক্ত উৎপাদক ব্যাকৰণ। উদাহৰণ—

Example

```
S → X a
X → a
X → aX
X → abc
X → ε
```

Type-1 ব্যাকৰণ আৰু অধিক সীমাবদ্ধ। এই Type-1 ব্যাকৰণে উৎপন্ন কৰে context-sensitive languages। ইয়াৰ উৎপাদন নিম্নোক্ত ৰূপত হ'ব লাগিব।

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

β

য'ত $A \in N$ (Non-terminal)

আৰু $\alpha, \beta, \gamma \in (T \cup N)^*$ (Strings of terminals and non-terminals)

α

যোগসূত্ৰ $()$ আৰু ϵ হ'ব লাগিব বিকৃত (empty) আৰু ϵ হ'ব লাগিব বিকৃতহীন।

(ব্যতিক্ৰমঃ $S \rightarrow \epsilon$ নিয়ম অনুমোদিত যদিহে সোঁফালে কোনো নিয়মত দেখা নাযায়।

ইয়াত বাওঁফালে কমেও এটাকৈ চিহ্ন ৰখাৰ অনুমতি দিয়ে আৰু সেইদৰে প্ৰসঙ্গও থাকিব। এই নিয়মত

γ (আলফা) আৰু β (বিটা)ৰ পৰিৱৰ্তন নহ'ব আৰু সেয়ে প্ৰসঙ্গ সংৰক্ষিত থাকে।

Type-0 ৰ দ্বাৰা উৎপন্ন (অসীমাবদ্ধ) ব্যাকৰণৰ ভাষা বেছি প্ৰকাশবাদী (কম সীমাবদ্ধতা)। ইয়াৰ উৎপাদনৰ ক্ষেত্ৰত কোনো সীমাবদ্ধতা নাই। Type-0 পৰা উৎপন্ন ভাষাক recursively enumerable language বোলা হয়। recursively enumerable language ৰ বাবে Turing machine ব্যৱহাৰ কৰা হয়।

ইয়াৰ উৎপাদন হ'ব পাৰে এই ৰূপত, $\alpha \rightarrow \beta$, য'ত α (আলফা) terminals আৰু nonterminals মিলি যোগসূত্ৰ। ইয়াত এটা হ'লৈও nonterminal থাকিব লাগিব আৰু ϵ (আলফা) null হ'ব নোৱাৰিব। β হ'ল এটা terminals আৰু nonterminals মাজৰ যোগসূত্ৰ।

Example

```
S → ACaB
Bc → acB
CB → DB
aD → Db
```

ব্যাকৰণ মূলতঃ নিয়মৰ সমষ্টি। প্ৰসঙ্গ মুক্ত ব্যাকৰণ হ'ল পুনৰাবৃত্তিৰে পুনৰ্লিখনৰ নিয়ম (বা উৎপাদন) যাক একেধৰণৰ আৰ্হি নিৰ্মাণৰ কাৰণে ব্যৱহাৰ কৰা হয়। প্ৰসঙ্গমুক্ত ব্যাকৰণ নিম্নলিখিত উপাদানসমূহেৰে গঠিত—

(ক) terminal চিহ্নৰ একোটা সমষ্টি যিবোৰ বৰ্ণমালাৰ আখৰ যিবোৰ ব্যাকৰণৰ দ্বাৰা উৎপন্ন একে নিয়মেৰে প্ৰকাশ পায়। যেনে— শব্দ

(খ) nonterminal চিহ্নৰ একোটা সমষ্টি, যি টাৰ্মিনেল চিহ্নৰ আৰ্হিৰ কাৰণে স্থানধাৰক যি নিৰ্মাণ হ'ব পাৰে nonterminal চিহ্নৰ দ্বাৰা। যেনে— বিশেষ্য খণ্ডবাক্য, ক্ৰিয়া খণ্ডবাক্য আৰু বাক্য।

(গ) উৎপাদনৰ একোটা সমষ্টি যিবোৰ একোটা নিয়ম nonterminal চিহ্নৰ স্থানান্তৰ বা পুনৰ্লিখনৰ (উৎপাদনৰ বাওঁ কাষে) লগত nonterminal চিহ্ন বা terminal চিহ্নৰ (উৎপাদনৰ সোঁ দিশে) মাজৰ যোগসূত্ৰ।

(ঘ) এক প্ৰাৰম্ভিক চিহ্ন, যি nonterminal চিহ্নৰ ৰূপত ব্যাকৰণৰ দ্বাৰা প্ৰাৰম্ভিক যোগসূত্ৰ স্থাপন কৰে।

প্ৰসঙ্গমুক্ত ব্যাকৰণৰ পৰা terminal চিহ্নৰ যোগসূত্ৰ স্থাপন কৰিবলৈ, আমি—

— প্ৰাৰম্ভিক চিহ্নৰ পৰা যোগসূত্ৰ আৰম্ভ কৰিব লাগিব; (Begin with a string consisting of the start symbol;)

— প্ৰাৰম্ভিক চিহ্নৰ প্ৰয়োগ কৰি বাওঁফালে উৎপাদন, প্ৰাৰম্ভিক চিহ্নৰ স্থান পৰিৱৰ্তন কৰি সোঁফালৰ উৎপাদন; (Apply one of the productions with the start symbol on the left hand side, replacing the start symbol with the right hand side of the production;)

যেতিয়ালৈকে সকলো nonterminal চিহ্ন terminal চিহ্নৰ দ্বাৰা প্ৰতিস্থাপিত নহয় তেতিয়ালৈকে nonterminal চিহ্ন নিৰ্বাচন কৰি প্ৰক্ৰিয়াটো দোহাৰিব লাগিব আৰু একদৰে সোঁফালে কিছুমান সংশ্লিষ্ট বা সম্পৰ্কিত উৎপাদন লৈ প্ৰক্ৰিয়াটোৰ পুনৰাবৃত্তি কৰিব লাগিব। (Repeat the process of selecting nonterminal symbols in the string, and replacing them with the right hand side of some corresponding production, until all nonterminals have been replaced by terminal symbols.)

প্ৰসঙ্গমুক্ত ব্যাকৰণ চাৰিটা উপাদানৰ সমষ্টি, $G = (V, \Sigma, P, S)$, য'ত

- V is a finite set of symbols called the *vocabulary* (or *set of grammar symbols*);
- $\Sigma \subseteq V$ is the set of *terminal symbols* (for short, *terminals*);
- $S \in (V - \Sigma)$ is a designated symbol called the *start symbol*;
- $P \subseteq (V - \Sigma) \times V^*$ is a finite set of *productions* (or *rewrite rules*, or *rules*).

V - হ'ল কেতবোৰ সীমিত চিহ্নৰ সমষ্টি শব্দ (বা ব্যাকৰণৰ চিহ্ন সমষ্টি);

$\Sigma \subseteq V$ -ক কোৱা হয় টাৰ্মিনেল চিহ্নৰ সমষ্টি (চমুকৈ টাৰ্মিনেল);

$S \in (V - \Sigma)$ হ'ল নিৰ্দিষ্ট বিশেষ চিহ্ন, যাক কোৱা হয় প্ৰাৰম্ভিক চিহ্ন;

$P \subseteq (V - \Sigma) \times V^*$ -ক কোৱা হয় সীমিত পৰিসৰৰ উৎপাদনৰ সমষ্টি (বা পুনৰ্লিখনৰ নিয়ম, বা নিয়ম)।

$N = V -$ -ক কোৱা হয় nonterminal চিহ্ন। সেইদৰে

আৰু প্ৰত্যেক উৎপাদন (A, ϵ) -কো উল্লেখ কৰা হয় $A \rightarrow \epsilon$ । A হ'ল $A \rightarrow \epsilon$ উৎপাদনৰ উৎপাদক নিয়মক কোৱা হয় epsilon rule, বা null rule।

ব্যাকৰণৰ নিয়মানুসৰি আমি ধাৰণা কৰি লওঁ যে আমাৰ দুই ধৰণৰ প্ৰতীক আছে—the terminals, ভাষাৰ অন্তৰ্গত বৰ্ণমালাৰ প্ৰতীক (the symbols of the alphabet underlying the languages) আৰু the nonterminals—এ terminal-বোৰৰ মাজত হোৱা সংযোগৰ জৰিয়তে variablesৰ দৰে আচৰণ কৰে (which behave like variables ranging over strings of terminals)। A হ'ল নিয়মৰ বিন্যাস $A \rightarrow \epsilon$ য'ত A হ'ল nonterminal একক। আনহাতে,

সোঁফালে থকা হ'ল terminal-ৰ একক আৰু বা nonterminals-ৰ প্ৰতীক। ইয়াতে আমি পোনপ্ৰথমে নিয় কৰি ল'ব লাগিব কৰ্ম (object) কি (কোনো প্ৰসঙ্গমুক্ত ব্যাকৰণত) আৰু তাৰ পাছত এইটো বুজিবলৈ চেষ্টা কৰা, কেনেকৈ ইয়াৰ ব্যৱহাৰ হৈছে। automataৰ বিপৰীতে ব্যাকৰণৰ ব্যৱহাৰ হয় চিনাক্ত কৰাৰ পৰিৱৰ্তে উৎপাদনৰ একক হিচাপেহে।

প্ৰায়োগিক ভাষাবিজ্ঞান

Wilkinsৰ মতে, প্ৰায়োগিক ভাষাবিজ্ঞান মানৱীয় ক্ষেত্ৰত ভাষাৰ ভূমিকা বুজাৰ সৈতে সম্পৰ্কিত আৰু তাৰদ্বাৰা ভাষা-সম্পৰ্কীয় সিদ্ধান্তৰ প্ৰয়োজন থকাসকললৈ সেই জ্ঞান আগবঢ়োৱা। প্ৰায়োগিক ভাষাবিজ্ঞান হ'ল আন্তঃসম্পৰ্কীয় বিষয় যিয়ে বাস্তৱ পৃথিৱীৰ ভাষা-সম্পৰ্কীয় সমস্যাৰ অনুসন্ধান আৰু সমাধান আগবঢ়ায়। দ্বিভাষিকতাবাদ, বহুভাষিকতাবাদ, কম্পিউটাৰ মাধ্যমেৰে যোগাযোগ, কথোপকথন বিশ্লেষণ, চিহ্ন ভাষাবিজ্ঞান, ভাষিক বিকাশ, বাচন বিশ্লেষণ, ভাষা শিক্ষা, ভাষা পৰিকল্পনা আৰু নীতি, প্ৰসঙ্গাৰ্থবিজ্ঞান, শব্দকোষ, অনুবাদ, ফৰেনছিক ভাষাবিজ্ঞান, ভাষিক নৃতত্ত্ব, স্নায়ুভাষাবিজ্ঞান, মনোভাষাবিজ্ঞান, সমাজভাষাবিজ্ঞান আৰু শিক্ষা বিষয় প্ৰধানকৈ প্ৰায়োগিক ভাষাবিজ্ঞানৰ অন্তৰ্গত বিষয়। ইয়াৰ উপৰি দ্বিতীয় ভাষাৰ শিক্ষণ আৰু শিকন এটা প্ৰধান শাখা হিচাপে সদায় গুৰুত্বপূৰ্ণ বুলি বিবেচিত হৈ আহিছে।

প্রাকৃতিক ভাষা প্রক্ৰিয়াকৰণ (Natural Language Processing) :

প্রাকৃতিক ভাষা প্রক্ৰিয়াকৰণ কি?

প্রাকৃতিক ভাষা প্রক্ৰিয়াকৰণ দৰাচলতে কম্পিউটাৰ ভাষাবিজ্ঞানৰ এক শাখাহে। মানৱ ভাষা আৰু কম্পিউটাৰ পৰস্পৰৰ মাজত প্ৰধানকৈ গুৰুত্ব দিয়া অধ্যয়ন ক্ষেত্ৰক প্রাকৃতিক ভাষা প্রক্ৰিয়াকৰণ বোলা হয়। মানৱ ভাষাক প্রক্ৰিয়াকৰণৰ জৰিয়তে কম্পিউটাৰৰ উপযোগীকৈ প্ৰস্তুত কৰাটোৱেই এই ক্ষেত্ৰৰ প্ৰধান লক্ষ্য। দৰাচলতে এন এল পি এনে এক মাধ্যম যি স্মাৰ্ট আৰু উপযোগী মাধ্যমেৰে মানৱ ভাষাৰ অৰ্থ-বিশ্লেষণ, বোধগম্য আৰু প্ৰাপ্ত কৰা এক উপায়বিশেষ। এন এল পিৰ মাধ্যমেৰে, উন্নীতকাৰকসকলে স্বয়ংক্ৰিয় সমীকৰণ, অনুবাদ, তথ্য অন্তৰ্ভুক্তিৰ চিনাক্তকৰণ, সম্পৰ্ক নিষ্কাশন, অনুভূতি বিশ্লেষণ, বাক্ চিনাক্তকৰণ আৰু বিষয়- বিভাগৰ দৰে কাৰ্য সম্পাদন কৰি জ্ঞান সংগঠিত বা গঠন কৰিব পাৰে।

এন এল পি (প্রাকৃতিক ভাষা প্রক্ৰিয়াকৰণ) মুখ্যতঃ এটা বহুমুখী বিষয় আৰু ভাষাবিজ্ঞানৰ লগত ঘনিষ্ঠভাৱে সম্পৰ্কিত। অৱশ্যে উল্লেখ কৰাটো উচিত হ'ব ভাষাবিজ্ঞানৰ সূত্ৰসমূহক সংগঠিত ৰূপত উপস্থাপনলৈকেহে ইয়াৰ পৰিধি। এন এল পি সংজ্ঞানাত্মক বিজ্ঞান, মনোবিজ্ঞান, দৰ্শন আৰু অংকশাস্ত্ৰ (বিশেষকৈ ইয়াৰ তৰ্কশাস্ত্ৰৰ দিশটো) আদি বিষয়ৰ লগতো গভীৰভাৱে সম্পৃক্ত।

কম্পিউটাৰ ভাষাবিজ্ঞানৰ শাখা হিচাপে ই আনুষ্ঠানিক ভাষা সিদ্ধান্ত, সংকলন প্ৰযুক্তি, প্ৰমেয়ৰ প্ৰামাণিকতা, যন্ত্ৰ শিকন আৰু মানৱ-কম্পিউটাৰৰ আন্তৰ্কাৰ্যৰ লগত সম্পৰ্কিত। ই 'কৃত্ৰিম বুদ্ধিমত্তা'ৰ লগতো সম্পৰ্কিত।

এন এল পিৰ বৰ্তমান প্ৰয়োগ ক্ষেত্ৰ প্ৰধানকৈ—

১. আখৰ জোঁটনি আৰু ব্যাকৰণ পৰীক্ষণ (spelling and grammar checking)
২. অপ্টিকেল অক্ষৰ স্বীকৃতি (optical character recognition (OCR))
৩. অন্ধ আৰু আংশিকভাৱে দেখা পোৱা লোকসলৰ বাবে পাঠৰ পঠন (screen readers for blind and partially sighted users)
৪. বৰ্ধিত আৰু বিকল্প যোগাযোগ; (অৰ্থাৎ বিশেষভাৱে সক্ষমলোকসলৰ বাবে যোগাযোগৰ ক্ষেত্ৰত থকা অসুবিধাসমূহ আঁতৰ কৰা)(augmentative and alternative communication (i.e., systems to aid people who have difficulty communicating because of disability))
৫. মেচিন বা যন্ত্ৰানুবাদত সহায় (যিয়ে মানৱ অনুবাদকাৰীক সহায় কৰা পদ্ধতিবিশেষ, যেনে—ভাষাবিশেষৰ খণ্ডবাক্যসমূহৰ অনুবাদৰ লগতে অনলাইন অভিধান আৰু ৱৰ্ড প্ৰছেছৰৰ লগত সংযুক্তকৰণ) (machine aided translation (i.e., systems which help a human translator, e.g., by storing translations of phrases and providing online dictionaries integrated with word processors, etc))
৬. অভিধান প্ৰণেতাৰ বাবে ব্যৱহাৰ্য আহিলা (lexicographers' tools)
৭. তথ্য আহৰণ (information retrieval)
৮. নথি বা তথ্য শ্ৰেণীবিভাগ (document classification)
৯. তথ্য নিষ্কাশন (text segmentation)
১০. প্ৰশ্ন-উত্তৰ (question answering)
১১. সংক্ষিপ্তকৰণ (summarization)
১২. পৰীক্ষাৰ মূল্যাংকন (exam marking)
১৩. পৰীক্ষাৰ মূল্যাংকন (exam marking)
১৪. প্ৰতিবেদন প্ৰস্তুতকৰণ (সম্ভৱত বহুভাষী) (report generation (possibly multilingual))
১৫. মেচিন অনুবাদ বা যন্ত্ৰানুবাদ (machine translation)
১৬. ইমেইল বোধগম্যতা (email understanding)
১৭. সংলাপ পদ্ধতি (dialogue systems)
১৮. (natural language interfaces to databases)

প্রাকৃতিক ভাষা প্রক্ৰিয়াকৰণৰ পদক্ষেপ (Steps in natural language processing)
ৰূপতাত্ত্বিক বিশ্লেষণ (Morphological Analysis)
বাক্যতাত্ত্বিক বিশ্লেষণ (Syntactic Analysis)
শব্দার্থ বিশ্লেষণ (Semantic Analysis)
বাচন বা কথা-বিশ্লেষণ (Discourse Analysis)
প্রসঙ্গার্থ বিশ্লেষণ (Pragmatic Analysis)।

প্ৰসঙ্গ মুক্ত ব্যাকৰণ (Context Free Grammar)—

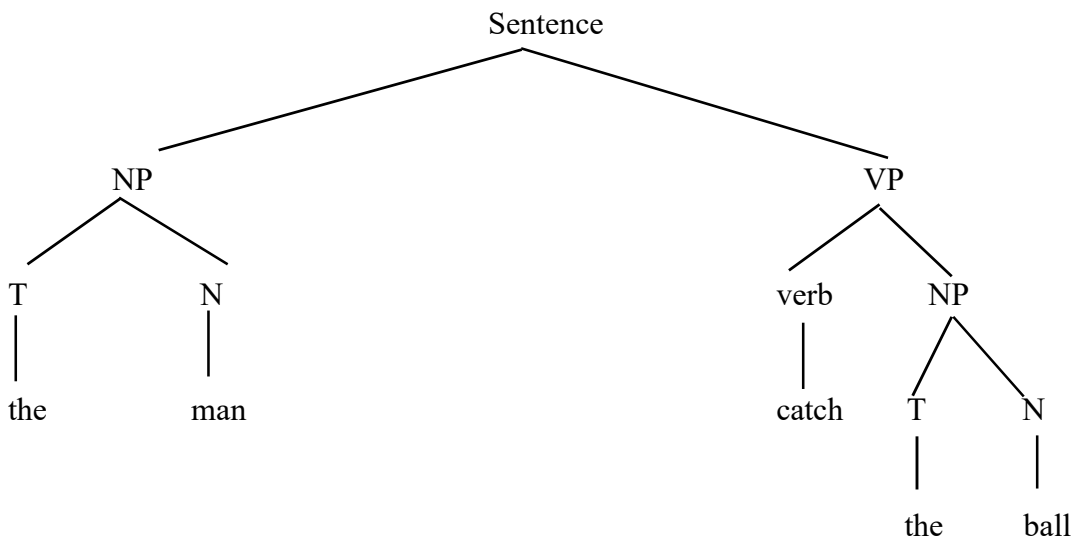
আনুষ্ঠানিক ভাষা প্ৰমেয় (Formal Language Theory) অনুসৰি প্ৰসঙ্গ মুক্ত ব্যাকৰণ হ'ল এক আনুষ্ঠানিক ব্যাকৰণ য'ত সকলোবোৰ উৎপাদনৰ সূত্ৰ $x \rightarrow y$ এই ৰূপত হয়। য'ত x হ'ল non-terminal চিহ্ন আৰু y হ'ল terminal-ৰ শব্দ-শৃংখল (string of terminals) অথবা non-terminal (y কেতিয়াবা শূন্যও (empty) হ'ব পাৰে)। ভাষাবিজ্ঞানত কোনো কোনোৱে প্ৰসঙ্গ মুক্ত ব্যাকৰণক পদগুচ্ছ সংগঠন ব্যাকৰণ (Phrase Structure Grammar) বুলিও অভিহিত কৰে। এইমতে বাক্য গঠন হয়—

- (i) Sentence \rightarrow NP + VP
- (ii) NP \rightarrow T + N
- (iii) VP \rightarrow Verb + N
- (iv) T \rightarrow The
- (v) N \rightarrow man, ball etc.
- (vi) Verb \rightarrow catch, hit etc.

ধৰা হ'ল, 'rewrite x as y' এই নিৰ্দেশানুসৰি আমি প্ৰত্যেকটি সূত্ৰক $x \rightarrow y$ হিচাপে ব্যাখ্যা কৰিলোঁ আৰু ফলত 'The man catch the ball' —এই বাক্যটিৰ ব্যুৎপত্তি দেখুৱাব পাৰোঁ। ব্যাকৰণৰ সূত্ৰ অনুসৰি ওপৰত দিয়া ধৰণে তলত ব্যুৎপত্তি দেখুওৱা হ'ল—

Sentence	
NP+VP	(i)
T+N+VP	(ii)
T+N+V+NP	(iii)
the +N+V+NP	(iv)
the + man +V+NP	(v)
the + man + catch +NP	(vi)
the + man + catch +T+N	(ii)
the + man + catch +the+N	(iv)
the + man + catch + the + ball	(v)

এই ব্যুৎপত্তি বৃক্ষৰেখাচিত্ৰ (Tree diagram) ৰ মাধ্যমেদি দেখুৱালে এনেধৰণৰ হ'ব—



আনুষ্ঠানিক ব্যাকৰণ প্ৰসংগমুক্ত হিচাপে বিৱৰ্চিত হ'ব পাৰে তেতিয়াহে যেতিয়া ইয়াৰ উৎপাদনৰ সূত্ৰসমূহ non-terminal-ৰ ক্ষেত্ৰতো নিৰ্বিশেষে ব্যৱহাৰ হ'ব পাৰে। কি চিহ্নেৰে পৰিৱেশিত হৈ আছে সেয়া বিচাৰ্য নহৈ বাওঁফালৰ পৰা একাদিক্ৰমে সোঁফাললৈ সমূহ স্থানান্তৰণ হ'ব। অনানুষ্ঠানিকভাৱে, প্ৰসংগমুক্ত ব্যাকৰণ এনে এক ব্যাকৰণ য'ত, যি কোনো বিন্দুত (point) যি কোনো পৰ্যায়ত non-terminal-ক বিন্ধিত কৰিব পৰা যায়। এই ব্যাকৰণ হ'ল কেতবোৰ strings of terminal-ৰ সমষ্টি যিবোৰৰ উৎপাদন start চিহ্নৰ পৰা আৰম্ভ হয়।

প্ৰসংগমুক্ত ব্যাকৰণ চাৰিটা উপাদানৰ দ্বাৰা গঠিত। সেয়া হ'ল—

$$G=(V, T, S, P)$$

G= Grammar

V= Set of variables or non terminal symbols.

T= Set of terminal symbols

S= Start symbols

P= production rule

(ক) terminal চিহ্ন হ'ল— একোটা সমষ্টি যিবোৰ বৰ্ণমালাৰ আখৰ যিবোৰ ব্যাকৰণৰ দ্বাৰা উৎপন্ন একে নিয়মেৰে প্ৰকাশ পায়। যেনে— শব্দ

(খ) nonterminal চিহ্নৰ একোটা সমষ্টি, যি টাৰ্মিনেল চিহ্নৰ আৰ্হিৰ কাৰণে স্থানধাৰক যি নিৰ্মাণ হ'ব পাৰে nonterminal চিহ্নৰ দ্বাৰা। যেনে— বিশেষ্য খণ্ডবাক্য, ক্ৰিয়া খণ্ডবাক্য আৰু বাক্য।

(গ) উৎপাদনৰ একোটা সমষ্টি যিবোৰ একোটা নিয়ম nonterminal চিহ্নৰ স্থানান্তৰ বা পুনৰলিখনৰ (উৎপাদনৰ বাওঁ কাষে) লগত nonterminal চিহ্ন বা terminal চিহ্নৰ (উৎপাদনৰ সোঁ দিশে) মাজৰ যোগসূত্ৰ।

(ঘ) এক প্ৰাৰম্ভিক চিহ্ন, যি nonterminal চিহ্নৰ ৰূপত ব্যাকৰণৰ দ্বাৰা প্ৰাৰম্ভিক যোগসূত্ৰ স্থাপন কৰে।

প্ৰসংগমুক্ত ব্যাকৰণৰ পৰা terminal চিহ্নৰ যোগসূত্ৰ স্থাপন কৰিবলৈ, আমি—

— প্ৰাৰম্ভিক চিহ্নৰ পৰা যোগসূত্ৰ আৰম্ভ কৰিব লাগিব; (Begin with a string consisting of the start symbol;)

— প্ৰাৰম্ভিক চিহ্নৰ প্ৰয়োগ কৰি বাওঁফালে উৎপাদন, প্ৰাৰম্ভিক চিহ্নৰ স্থান পৰিৱৰ্তন কৰি সোঁফালৰ উৎপাদন; (Apply one of the productions with the start symbol on the left hand side, replacing the start symbol with the right hand side of the production;)

যেতিয়ালৈকে সকলো nonterminal চিহ্ন terminal চিহ্নৰ দ্বাৰা প্ৰতিস্থাপিত নহয় তেতিয়ালৈকে nonterminal চিহ্ন নিৰ্বাচন কৰি প্ৰক্ৰিয়াটো দোহাৰিব লাগিব আৰু একদৰে সোঁফালে কিছুমান সংশ্লিষ্ট বা সম্পৰ্কিত উৎপাদন লৈ প্ৰক্ৰিয়াটোৰ পুনৰাবৃত্তি কৰিব লাগিব। (Repeat the process of selecting nonterminal symbols in the string, and replacing them with the right hand side of some corresponding production, until all nonterminals have been replaced by terminal symbols.)

তলত উদাহৰণেৰে সৈতে প্ৰসংগমুক্ত ব্যাকৰণৰ ধাৰণাটো উল্লেখ কৰা হ'ল—

$$N = \{S\}, T = \{a, b\}, \text{ Starting symbol} = S, P = S \rightarrow SS \mid aSb \mid \varepsilon$$

উপৰিউক্ত উৎপাদন বিধি অনুযায়ী আমি এনেধৰণৰ উৎপাদন এটা পাব পাৰো—

$$S > SS \ (S \rightarrow aSb)$$

$$S > aSbS \ (S \rightarrow \varepsilon)$$

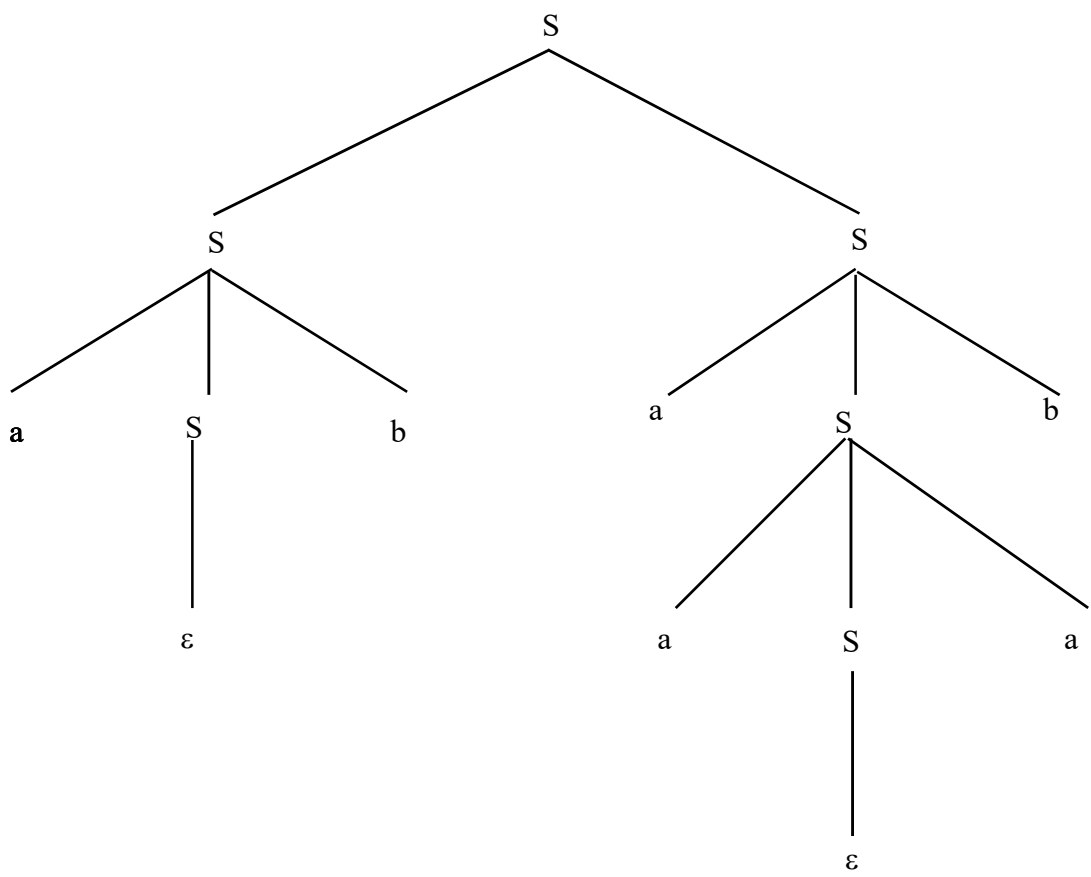
$$S > abS$$

$$S > abaSb \ (S \rightarrow aSb)$$

$$S > abaaSbb \ (S \rightarrow aSb)$$

$$s > abaabb \ (S \rightarrow \varepsilon)$$

ইয়াক বৃক্ষ চিত্ৰে দেখুৱালে এনেধৰণৰ হ'ব—



প্ৰসঙ্গমুক্ত ব্যাকৰণত বাক্য বিশ্লেষণ (Parsing in CFG)

স্বাভাৱিক ভাষা প্ৰক্ৰিয়াকৰণৰ প্ৰসঙ্গত, Parsing-এ কোনো প্ৰদত্ত বাক্যক স্বতঃস্ফূৰ্তভাৱে বিশ্লেষণৰ প্ৰক্ৰিয়াক বুজায়। বাক্যৰ সম্ভাৱ্য অন্তৰ্নিহিত গঠনত শব্দ-ক্ৰম অৱলোকন কৰে। ইয়াক এনেকৈয়ো ক'ব পাৰি— কোনো এক পাঠৰ বাক্য গঠনৰ প্ৰক্ৰিয়াটো নিৰ্দিষ্ট শব্দক্ৰমৰ জৰিয়তে কিদৰে গঠন হৈছে সেয়া ভাষাটোৰ অন্তৰ্নিহিত ব্যাকৰণৰ ভিত্তিত বিশ্লেষণ কৰে।

Simply speaking parsing in NLP is the process of determining the syntactic structure of a text by analyzing its constituent words based on an underlying grammar (of the language)

Parsing-ৰ কাৰণে প্ৰয়োজন হয় প্ৰদত্ত ভাষাটোৰ গাণিতিক আৰ্হি একোটাৰ। এটা আনুষ্ঠানিক ব্যাকৰণ গঠন হয় নিয়মৰ সমষ্টিৰ দ্বাৰা, যিয়ে প্ৰদত্ত ভাষাটোৰ উপাদানসমূহ বৰ্ণনা কৰে। উদাহৰণস্বৰূপে আমি শব্দৰ কথা ক'ব পাৰোঁ, বাক্য গঠনৰ সংযুক্তি অথবা বাক্য সংঘটনৰ প্ৰক্ৰিয়া হ'ব পাৰে। নিয়মসমূহ সম্পূৰ্ণভাৱে বাক্যগত অথবা ব্যাকৰণগত বিষয়ৰ হ'ব পাৰে, যেনে— ব্যাকৰণগত কাৰ্য (Grammatical function), কৰ্তা-ক্ৰিয়াৰ মাজত থকা বিধি, অথবা শব্দ গঠনৰ ক্ৰম ইত্যাদি।

Parsing তিনিটা উপাদানৰ ওপৰত নিৰ্ভৰশীল—

- (ক) শব্দ (Lexicon)
- (খ) শ্ৰেণীকৰণ (Categorization)
- (গ) ব্যাকৰণ বিধি (Grammar rules)

প্ৰসঙ্গমুক্ত ব্যাকৰণত Parsing-ৰ জৰিয়তে শব্দৰ শ্ৰেণীবিভাজন (যেনে— বিশেষ্য, বিশেষণ, সৰ্বনাম আদি) নিৰ্ণয় কৰি বাক্যসমূহ ব্যাকৰণগত হয়নে নহয় সেইটো পৰীক্ষা কৰিব পাৰি। এনে কৰিবলৈ সাধাৰণতে দুটা উপায় অৱলম্বণ কৰা হয়।

- (ক) Top-down parsing
- (ক) Bottom-up parsing

তলত প্ৰসঙ্গমুক্ত ব্যাকৰণৰ এটা উদাহৰণ দাঙি ধৰি কিদৰে বাক্য বিশ্লেষণ কৰা হয়, দেখুওৱা হ'ল—

G= (V, T, S, P)
V= that, this, a, the, man, book, flight, meal
T= S, NP, NOM, VP, Det., Noun, Verb, Aux.)
S= Start symbols
P= production rule

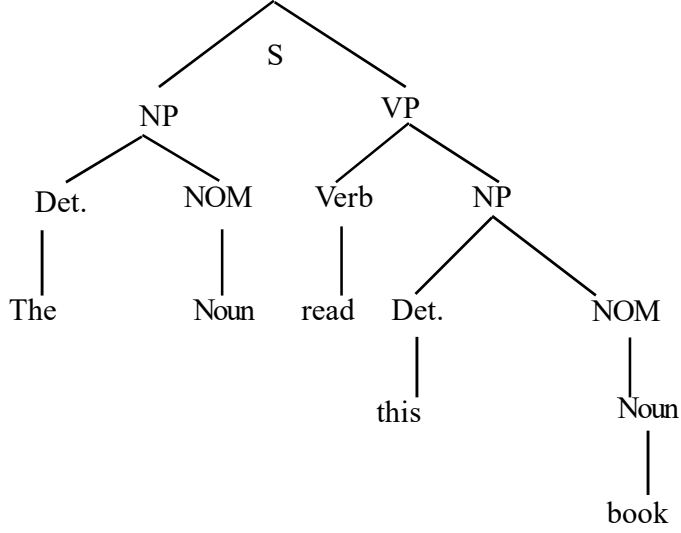
S→NP VP	Det → that/ this/ a/ the
S→ Aux. NP VP	Noun →book/ flight/ meal/ man
S→ VP	Verb→ book/ include/ read
NP→ Det NOM	Aux.→ does
NOM→ Noun	
NOM →Noun NOM	
VP→ Verb	
VP→Verb NP	

পুনৰ্লিখন-বিধি অনুযায়ী ব্যাকৰণৰ প্ৰয়োগ

S → NP VP
→Det NOM VP
→The NOM VP
→The Noun VP
→The man VP
→The man Verb NP

- The man read NP
- The man read Det NOM
- The man read this NOM
- The man read this Noun
- The man read this book.

বৃক্ষ চিত্ৰৰ দ্বাৰা বাক্যতাত্ত্বিক বিশ্লেষণ কৰিলে ই এনেধৰণৰ হ'ব—



পুনৰ্নিৰ্মাণ প্ৰক্ৰিয়াৰ দ্বাৰা আমি এনেধৰণে অসংখ্য বাক্য সৃষ্টি কৰি থাকিব পাৰোঁ। উল্লেখযোগ্য কথা হ'ল—
 প্ৰসঙ্গমুক্ত ব্যাকৰণে আনুষ্ঠানিক ভাষাক সংজ্ঞায়িত কৰে যি বাক্যৰ সকলো শব্দ-গুচ্ছ (string), যিবোৰ ভাষা একোটাৰ
 ব্যাকৰণ সন্মত ৰীতি অনুযায়ী উৎপাদন হয়। ব্যাকৰণৰ নিয়মানুযায়ী গঠিত এনেবোৰ বাক্য ব্যাকৰণগত। নিয়মৰ আওতাৰ
 বাহিৰত থকাবোৰক কোৱা হ'ব অব্যাকৰণগত।