কম্পিউটাৰ ভাষাবিজ্ঞান

কম্পিউটাৰ ভাষাবিজ্ঞান হ'ল ভাষাবিজ্ঞান আৰু কম্পিউটাৰ বিজ্ঞানৰ সংযুতিত গঢ়ি উঠা আন্তৰ্শৃংখলিত 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—

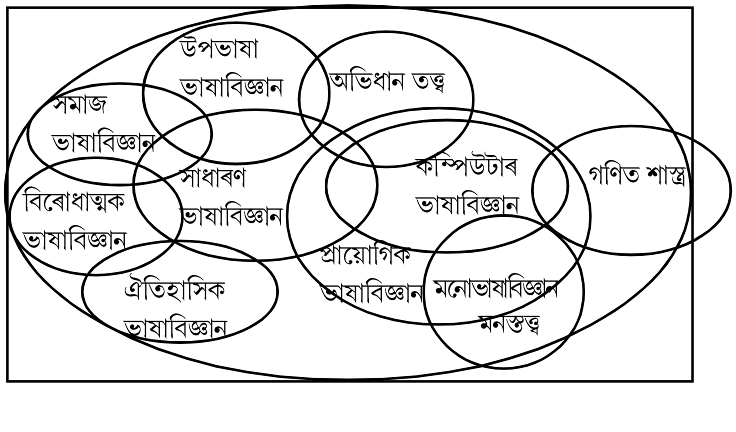
যন্ত্ৰানুবাদ (Machine translation)

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 (Transformal 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 সিবোৰক ব্যৱহাৰ কৰিছে। এই কাৰ্যক্ৰমবোৰৰ মূল্যাংকন 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 (appliscations) 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 ছফ্‌টৱেৰ 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 সমস্যাই। 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, যথাযথ শব্দৰ প্ৰয়োগ (word sense disembiguation), শব্দকোষ আৰু পাঠ নিৰ্মাণ, আৰু এটা ভাষাৰ পৰা আন এটা ভাষালৈ স্বতস্ফুতঃ যন্ত্ৰানুবাদ আদি।

কম্পিউটাৰ বিজ্ঞানৰ আৰম্ভণি

কম্পিউটাৰ ভাষাবিজ্ঞানৰ আৰম্ভণি হয় আমেৰিকাত ১৯৫০ চনৰ পৰা। ৰাছিয়াৰ বিজ্ঞান গৱেষণা পত্ৰিকা কম্পিউটাৰৰ দ্বাৰা স্বতস্ফূৰ্তঃ 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 উদ্দেশ্যতেই 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 অৰ্ন্তনিহিত জ্ঞান (Implicit encoding of linguistic knowledge)

(খ) পৰিসংখ্যন পদ্ধতি বা যন্ত্ৰ-শিকন পদ্ধতিৰ প্ৰয়োগ

কম্পিউটাৰ ভাষাবিজ্ঞানৰ ক্ষেত্ৰ

কম্পিউটাৰ ভাষাবিজ্ঞানৰ ক্ষেত্ৰসমূহ প্ৰধানকৈ—

(ক) মেছিন অনুবাদ বা যন্ত্ৰানুবাদ (Machine Translation)

(খ) বাক্-স্বীকৃতকৰণ (speech recognition)

(গ) মানুহ-যন্ত্ৰৰ পাৰস্পৰিক প্ৰদান প্ৰদান (man-machine interfaces)

(ঘ) ব্যাকৰণ শুদ্ধিকৰণ (Grammar Checker)

(ঘ) 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।