



Vol. 40 | No. 3 | 1997



Check for updates

# সাহিত্য পত্রিকা

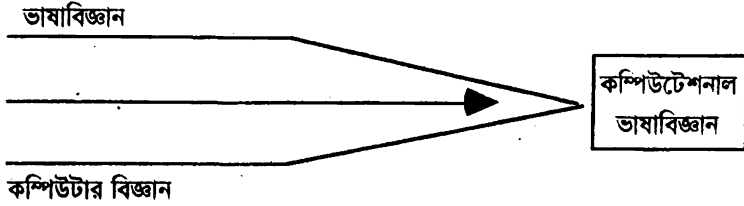
journal.bangla.du.ac.bd

কম্পিউটেশনাল ভাষাবিজ্ঞানের ভূমিকা

|                              |  |
|------------------------------|--|
| Volume                       | 40   |
| Issue                        | 3  |
| Year                         | 1997   |
| ISSN                         | 0558-1583  |
| eISSN                        | 3006-886X  |
| Author(s)                    | Hakim Arif   |
| Published online             | February 1, 1997   |
| DOI                          | 10.62328/sp.v40i3.10.1   |
| Link to article              | <a href="https://doi.org/10.62328/sp.v40i3.10.1">https://doi.org/10.62328/<br/>sp.v40i3.10.1</a> |
| Pages                        | 193-204  |
| Publisher                    | University of Dhaka  |
| Copyright                    | সাহিত্য পত্রিকা  |
| Designed and<br>Developed by | Zobayer Abdullah   |

পৃথিবীর মানবীয় ভাষাসমূহের চারটি ভাষিক স্তর যথা-ধ্বনি, রূপ, বাক্য ও অর্থের বিজ্ঞানসম্মত বিশ্লেষণের উপর ভিত্তি করেই মানববিদ্যারূপে গড়ে উঠেছে ভাষাবিজ্ঞান। কারণ ভাষাবিজ্ঞানের এই বিকাশ প্রতিটি দেশেই দেশীয় ভাষা এবং সাহিত্যকে কেন্দ্র করে পূর্ণতা পেয়েছে। পরবর্তীকালে সমাজবিজ্ঞান, নৃবিজ্ঞান অথবা মনোবিজ্ঞান ভাষাকে গুরুত্বপূর্ণ প্রপঞ্চ গণ্য করে সংশ্লিষ্ট দৃষ্টিকোণ থেকে ভাষাচর্চায় এগিয়ে আসে। এভাবে ভাষা গবেষণায়, ভাষার বিভিন্ন উপাত্ত বর্ণনায় সামাজিক বিজ্ঞানের রীতি ও পদ্ধতির প্রয়োগ হওয়ার ফলে কালক্রমে ভাষাবিজ্ঞান সামাজিক বিজ্ঞানের বিষয়ে পরিণত হয় এবং ভাষাবিজ্ঞানে বহুতর ‘সহ-শাখা’র জন্মলাভ করে। আবার বিশুদ্ধ বিজ্ঞান তথা প্রযুক্তি বর্তমান বিশ্বে একটি সর্বতোমুখী শাস্ত্র। ফলে সামাজিক বিজ্ঞান বা মানববিদ্যার বিষয় হিসেবে ভাষাবিজ্ঞানের নিজস্ব অঙ্গনেও আজকাল বিশুদ্ধ বিজ্ঞানের অবাধ আনাগোনা।<sup>১</sup> আর বিশুদ্ধ বিজ্ঞানের প্রযুক্তি নির্ভর গুরুত্বপূর্ণ শাস্ত্র (ফলিত) কম্পিউটার বিজ্ঞান এবং ভাষাবিজ্ঞানের সমন্বয়ে গড়ে উঠেছে ভাষাবিজ্ঞান অথবা কম্পিউটার বিজ্ঞানের সম্ভাবনাপূর্ণ সহ-শাস্ত্র কম্পিউটেশনাল ভাষাবিজ্ঞান।

কম্পিউটেশনাল ভাষাবিজ্ঞান মূলত মানবীয় বা প্রাকৃত ভাষার (natural language) বিভিন্ন স্তর যেমন ধ্বনি, রূপ, বাক্য ও অর্থের সংগঠনের কম্পিউটারায়ন বা কম্পিউটার প্রক্রিয়াকরণ প্রণালীর শাস্ত্র।<sup>২</sup> তাই এটি কম্পিউটার বিজ্ঞানের তত্ত্ব বা কম্পিউটারের বিভিন্ন কাঠামো বা সংগঠনের বর্ণনা নয়, বরং মানবীয় ভাষার বিভিন্ন স্তরে কম্পিউটারায়নে ক্রিভাবে উপযোগী করে তোলা যায় এবং কম্পিউটারায়নের জন্য ঐসকল স্তরের সূত্রাবদ্ধ ও যৌক্তিক ক্রমবিন্যস্ত পদ্ধতি উদ্ভাবন করা যায়, তারই অনুপঞ্জ বিশ্লেষণ। ফলে এ শাস্ত্রে কম্পিউটার বিজ্ঞানের চেয়ে ভাষাবিজ্ঞানে অধিকতর গুরুত্ব আরোপ করা হয়। তবে, কম্পিউটেশনাল ভাষাবিজ্ঞানের ভিত্তি হচ্ছে কম্পিউটার। কারণ কম্পিউটার ইনপুট করার উদ্দেশ্যেই বিভিন্ন ভাষিক স্তর বা উপাদানের ক্রমবিন্যস্ত পদ্ধতি বা আলগোরিদম (algorithm) প্রণয়ন কার্যক্রম পরিচালিত ও আবর্তিত। সর্বোপরি কম্পিউটেশনাল ভাষাবিজ্ঞান হচ্ছে ভাষাবিজ্ঞান ও কম্পিউটার বিজ্ঞানের একটি আন্তঃশাস্ত্রীয় নির্মিতি। নিম্নের চিত্রে এটি ব্যাখ্যায়-



কম্পিউটেশনাল ভাষাবিজ্ঞানের মূল লক্ষ্য ও পরিণতি হচ্ছে যান্ত্রিক অনুবাদ (machine translation)। কারণ প্রাথমিক পর্যায়ে যান্ত্রিক অনুবাদ কার্যক্রমে মানবীয় ভাষার যন্ত্রবোধ্য ভাষাতাত্ত্বিক সূত্রপরম্পরার শূন্যতার কারণেই কম্পিউটেশনাল ভাষাবিজ্ঞান ধারণাটি দানা বাঁধতে শুরু করে। সেই বিবেচনায় এই শাস্ত্রটির গোড়াপত্তন হয় ১৯৩০ খ্রিষ্টাব্দ সনের দিকে।<sup>৩</sup> তবে দ্বিতীয় বিশ্বযুদ্ধের পরে যুক্তরাষ্ট্র, যুক্তরাজ্য, সোভিয়েত ইউনিয়ন, জাপান, চীন, ইসরাইল, মেক্সিকো প্রভৃতি দেশে যান্ত্রিক অনুবাদের কয়েকটি প্রকল্প যখন ভাষাতাত্ত্বিক ব্যর্থতার কারণে পরিত্যক্ত হয় তখন থেকেই মূলত ইউরোপ ও আমেরিকার বিভিন্ন বিশ্ববিদ্যালয়ে কম্পিউটেশনাল ভাষাবিজ্ঞানের শাস্ত্র-ভিত্তিক অধ্যয়ন ও গবেষণা শুরু হয়। কম্পিউটেশনাল ভাষাবিজ্ঞানের আদিপুরুষ হিসেবে যান্ত্রিক অনুবাদের জনক ওয়ারেন ওয়েভারকেই অভিহিত করা যায়।<sup>৪</sup>

যান্ত্রিক অনুবাদ ব্যতীত কম্পিউটেশনাল ভাষাবিজ্ঞান প্রধানত মানবীয় বা প্রাকৃত ভাষার রূপ, কাঠামো, সংগঠন এবং সাহিত্য-উপাদানের কম্পিউটারায়ন সংক্রান্ত যে সব গবেষণাকর্মকে উপজীব্য করে তাদের মধ্যে উল্লেখযোগ্য কয়েকটি<sup>৫</sup> হচ্ছে—

- ১.০ মানবীয় ভাষার ভাষা-উপাস্তের বিশ্লেষণ সংগ্রহ ও বড় ধরনের পরিসংখ্যানগত মূল্যায়নের জন্য কার্যক্রম প্রণয়ন। যেমন—
- ১.১ ভাষার শব্দ সাজানো ও মুদ্রণ
- ১.২ লেখক ব্যবহৃত শব্দাবলী বা বিষয়সমূহের বর্ণানুক্রেমিক সূচি তৈরি
- ১.৩ সংশ্লিষ্ট গ্রন্থ বা নির্দিষ্ট যুগের অথবা কোন লেখকের রচনায় ব্যবহৃত শব্দের পৌনঃপুনিকতা নির্ধারণ
- ১.৪ ভাষার বিভিন্ন উপাদান যেমন— ধ্বনি, শব্দ, বর্ণ, ইত্যাদি পৌনঃপুনিকতা যাচাই
- ১.৫ সুনির্ধারিত মুখ্য শব্দ (key-words) দিয়ে স্বয়ংক্রিয়ভাবে নির্দিষ্ট Text-এর বর্ণানুক্রেমিক তালিকা/নির্ঘণ্ট তৈরি
- ২.০ ভাষিক আচরণ অনুকরণের জন্য ভাষা-মডেল প্রক্রিয়াকরণ

৩.০ কম্পিউটার কর্তৃক ব্যাখ্যাযোগ্য ভাষিক জ্ঞান বা মডেলসমূহ যথাযথভাবে তুলে ধরার জন্য formalism-এর উন্নয়ন।

৪.০ বাচন প্রক্রিয়াকরণ। ইত্যাদি

যান্ত্রিক অনুবাদ (machine translation) : কম্পিউটেশনাল ভাষাবিজ্ঞানের কেন্দ্রবিন্দু হচ্ছে যান্ত্রিক অনুবাদ। আবার যান্ত্রিক অনুবাদ কার্যক্রমের সফলতার উদ্দেশ্যেই যেহেতু কম্পিউটেশনাল ভাষাবিজ্ঞানের যাত্রা শুরু হয়েছে তাই যান্ত্রিক অনুবাদ এ-শাস্ত্রের ভিত্তিভূমি হিসেবেও গুরুত্ববাহী। যান্ত্রিক অনুবাদ বলতে প্রধানত একটি মানবীয় ভাষাকে অন্য মানবীয় ভাষায় অনুবাদ করার ক্ষেত্রে কম্পিউটারের অংশগ্রহণকেই বুঝানো হয়। অর্থাৎ সুনির্দিষ্ট কম্পিউটার প্রোগ্রামের সহায়তায় একটি মানবীয় ভাষার নির্ধারিত ট্রেসটকে অন্য একটি মানবীয় ভাষার অনুরূপ ট্রেসটে প্রেরণ করাই হচ্ছে যান্ত্রিক অনুবাদ।<sup>৬</sup>

কম্পিউটার বর্তমান বিশ্বের সর্বশেষ প্রযুক্তি-নির্ভর যান্ত্রিক আবিষ্কার এবং পৃথিবীর সমস্ত যন্ত্রের তুলনায় অধিক সুবেশক। কারণ নির্দিষ্ট যান্ত্রিক কাঠামোর মধ্যেই শুধু প্রোগ্রাম পরিবর্তন করে কম্পিউটারে যত বৈচিত্র্যধর্মী ক্রিয়া সম্পাদিত হয় তা অন্য কোন যন্ত্রে সম্ভব নয়। আর অনুবাদ যেহেতু একটি জটিল, শ্রম ও সময় সাপেক্ষ ব্যাপার, তাই কম্পিউটার আবিষ্কারের পরেই মানুষ অনুবাদ প্রক্রিয়ার কম্পিউটারায়নে উৎসাহী হয়ে উঠে এবং ১৯৩০ খ্রিষ্টাব্দে সনে শুরু হয় প্রথম যান্ত্রিক অনুবাদ কার্যক্রম। কিন্তু অনুকূল যান্ত্রিক পরিবেশ সৃষ্টি না হওয়ার কারণে তা ব্যর্থতায় পর্যবসিত হয়। দ্বিতীয় বিশ্বযুদ্ধোত্তর কালে তথ্য প্রবাহের অভাবনীয় উন্নতি, ইলেকট্রনিক কম্পিউটারের আবিষ্কার এবং উচ্চতর কোড-ব্রেকিং কৌশলের সফলতায় নতুন করে যান্ত্রিক অনুবাদ কার্যক্রম শুরু হয়। কিন্তু কম্পিউটার বোধ্যক্রমবিন্যস্ত নির্ভুল ভাষাতাত্ত্বিক সূত্র পরম্পরা উদ্ভাবিত না হওয়ার কারণে তাও সফলতায় উন্নীত হয়নি। এ পর্যায়ে বিজ্ঞানীরা মানবীয় ভাষার শব্দ-সংগঠন, বাক্য সংগঠন ও অর্থ পরম্পরাকে উপেক্ষা করে শুধু কতিপয় শব্দের সমষ্টি হিসেবে গণ্য করে যান্ত্রিক অনুবাদে অগ্রবর্তী হয়েছিলেন।<sup>৭</sup> ফলে এক্ষেত্রে ব্যর্থতা ছিল অনিবার্য।

কোন মানবীয় ভাষাই শুধু কতিপয় শব্দের সমষ্টি মাত্র নয়। প্রত্যেকটি প্রাকৃত বা মানবীয় ভাষারই রয়েছে একটি নিজস্ব ধ্বনি শৃঙ্খলা, শব্দ সংগঠন, বাক্য সংগঠন এবং অর্থপরম্পরা। তাছাড়া ভাষার শব্দসমূহের অর্থে বিদ্যমান আছে নানারকম দ্ব্যর্থকতা, বিভ্রমতা এবং সংশ্লিষ্ট সংস্কৃতি ও আচার নির্ভরতা। প্রকৃতপক্ষে মানবীয় ভাষা তার সহজাত বৈশিষ্ট্যের কারণেই দ্ব্যর্থক, অস্পষ্ট ও অযথার্থ।<sup>৮</sup> অধ্যাপক উদর নারায়ণ সিংহ (কার্তিক ১৪০৩ : ৮৩) মানবীয় ভাষার বৈশিষ্ট্য প্রসঙ্গে বলেছেন—

“ভাষার পৃথিবীতে একটা আপাত নৈরাজ্য আছে যার প্রথম দর্শনে যেকোনো তাত্ত্বিক বলবেন

- ক. ভাষা অত্যন্ত বিভেদ-বিভ্রমী;
- খ. অত্যন্ত ব্যক্তিগত, অথচ একই সঙ্গে সমষ্টির সংজ্ঞায় বাঁধা;
- গ. অতীব ভঙ্গুর ও মরণশীল;
- ঘ. জাতি, ধর্ম, শ্রেণী, লিঙ্গ প্রভৃতি বিভিন্ন সামাজিক এককের সঙ্গে সম্পর্কিত অথচ একেবারে স্বতন্ত্র; এবং
- ঙ. ভাষার আকরণই এমন যে তার এক পা পৃথিবীতে, অন্যটি আকাশে— একপক্ষে ধ্বনি নামক ভাষাজটি যেমন বাস্তব উচ্চারণ এবং ব্যবহার্য সত্য, অর্থের দিকটি যেন তেমনি না জানা না বোঝা না ধরা এক উচ্চারণ বিমূর্ত কল্পনা।”

মানবীয় ভাষার বিভিন্ন উপাদান তথা অর্থের এই যে অপরিমেয় বিমূর্ততা, দ্ব্যর্থকতা, অস্পষ্টতা তা যান্ত্রিক অনুবাদ কালে বিজ্ঞানীদের বিভ্রান্ত ও কিংকর্তব্যবিমূঢ় করেছে। ফলশ্রুতিতে তা নিরসনেই শুরু হয়েছে কম্পিউটেশনাল ভাষাবিজ্ঞান ভিত্তিক কম্পিউটার বিজ্ঞানে দক্ষ ও অভিজ্ঞ ভাষাবিজ্ঞানীর সহায়তায় কম্পিউটার বোধ্য মানবভাষা বিন্যস্ত করার কার্যক্রম।

কম্পিউটার প্রোগ্রামিং তথা পরিচালনার জন্য একমাত্র ও প্রধান মাধ্যম হচ্ছে কম্পিউটার নিয়ন্ত্রিত এক বিশেষ যান্ত্রিক ভাষা। যেমন —COBOL, ALGOL, FORTRAN, PASCAL ইত্যাদি। তবে কম্পিউটারের এই বিশেষ যান্ত্রিক ভাষা কম্পিউটার-বোধ্য, যুগ্মরীতির আঙ্গিক ও যুক্তি নির্ভর। কিন্তু মানবীয় বা প্রাকৃত ভাষার ক্ষেত্রে উয় যেহেতু নান্দায়কম দ্ব্যর্থকতা ও অস্পষ্টতায় পরিপূর্ণ তাই এ ভাষা কম্পিউটারের কাছে অবোধ্য ও জটিল প্রতীয়মান হয়।<sup>১৯</sup> আর যান্ত্রিক অনুবাদে যেহেতু অনুবাদ প্রক্রিয়া যন্ত্র বা কম্পিউটারের অভ্যন্তরেই সংগঠিত হয় তাই এক্ষেত্রে যে দুটি ভাষায় অনুবাদ কার্যক্রম আবর্তিত হবে সে দুটি ভাষা অর্থাৎ উৎসভাষা (source language) এবং লক্ষ্যভাষাকে (target language) অবশ্যই কম্পিউটারের কৃত্রিম ভাষার মত কম্পিউটার-বোধ্য, যুগ্মরীতির আঙ্গিক ও যুক্তিনির্ভর হতে হবে। যে প্রক্রিয়ায় মানবীয় ভাষাসমূহকে এই স্তরে উন্নীত করা হয় সেই প্রক্রিয়াকে কম্পিউটারের পরিভাষায় আলগোরিদম (algorithm) বলা হয়।

যে কোন নিয়মাবদ্ধ পদ্ধতিকে কিছু নির্দেশের পর্যায়ক্রমিক বিন্যাসে বিন্যস্ত করা যায়। এই বিন্যাসকে বলা হয় ঐ পদ্ধতির algorithm।<sup>১০</sup> অর্থাৎ

স্বয়ংক্রিয়ভাবে সমস্যা সমাধানের জন্য গাণিতিক বা যৌক্তিক প্রক্রিয়ায় বিন্যাসকৃত নির্ভুল ও কম্পিউটার-বোধ্য নিয়মাবলীকেই প্রধানত কম্পিউটারের আলগোরিদম বলে গণ্য করা যেতে পারে।<sup>১১</sup> এই আলগোরিদমকে 'প্রাধাম' বা 'সমাধানসূচক লজিক সূত্রের সেট' বলেও কেউ কেউ অভিহিত করেছেন।<sup>১২</sup> সর্বোপরি যন্ত্র বা কম্পিউটার কর্তৃক প্রক্রিয়াকরণ বা সম্পাদন সমর্থ ক্রমবিন্যস্ত নিয়মাবলীই হচ্ছে আলগোরিদম। যন্ত্র বা কম্পিউটারের এই পরিভাষাটি যে কোন সাধারণ দ্রব্য বা বস্তুর নিয়মান্বয়ের ক্ষেত্রেও প্রযুক্ত হতে পারে। যেমন- কফি তৈরির ক্ষেত্রে 'জল ফুটানো থেকে শুরু করে কফি তৈরি পর্যন্ত' কার্যপ্রণালিটি যদি নিয়মাবদ্ধ বা ক্রমবিন্যস্ত করা হয়, তবে তা হবে কফি তৈরির আলগোরিদম।<sup>১৩</sup>

যান্ত্রিক অনুবাদের ক্ষেত্রে মানবীয় ভাষা তথা উৎস ও লক্ষ্য ভাষার আলগোরিদম তৈরি অতিশয় জরুরী। কারণ মানবীয় ভাষাসমূহের আলগোরিদম তৈরির কাজ সম্পন্ন হওয়ার পরেই কেবল তারা কম্পিউটারের মধ্যবর্তী বা যান্ত্রিক ভাষার বোধগম্য হবে, যুগ্মরীতির আঙ্কিক নিয়মানুসারী ও যৌক্তিক পরম্পরায় বিশ্লেষিত হবে এবং কম্পিউটারায়নে উপযোগী বলে প্রতীয়মান হবে। মানবীয় ভাষা তথা উৎস ও লক্ষ্যভাষাদুটির আলগোরিদম যখন তৈরি হবে অর্থাৎ কম্পিউটার বোধ্য হওয়ার পাশাপাশি বিভিন্ন ভাষিক স্তর যেমন ধ্বনি, রূপ, বাক্য ইত্যাদির সংগঠন ও অর্থের সম্পৃক্ততা, দ্ব্যর্থকতা কেটে যাবে এবং গাণিতিক নিয়ম ও যুক্তি-সূত্রের অনুগামী হবে তখন এদেরকে বলা হবে 'ফরমাল ভাষা' (formal language)। কারণ যে কোন মানবীয় ভাষার বিভিন্ন ভাষিক স্তরের ভাষাতাত্ত্বিক সূত্র পরম্পরা যা গাণিতিক নিয়ম ও যুক্তি-সূত্রের অনুসারী এবং সম্পৃক্ততা ও নির্ভুলতার জন্য বিশিষ্ট, তাকেই 'ফরমাল ভাষা' বলা হয়ে থাকে।<sup>১৪</sup>

যান্ত্রিক অনুবাদের উপপ্রমেয় হিসেবে যদি বাংলা ও ইংরেজি ভাষাকে বিবেচনায় আনি, অর্থাৎ বাংলাকে উৎস ভাষা ও ইংরেজিকে লক্ষ্যভাষা ধরে যান্ত্রিক অনুবাদ কার্যক্রম শুরু করে তবে প্রথমেই এ দুটি মানবীয় ভাষাকে ফরমাল ভাষায় উন্নীত করতে হবে। বাংলা ও ইংরেজিকে ফরমাল ভাষায় উন্নীত করতে পারলেই এ দুটি ভাষার যান্ত্রিক অনুবাদ সহায়ক আলগোরিদম তৈরী হবে বলে ধরে নেওয়া যায়। এক্ষেত্রে প্রথম বিবেচ্য হল, বাংলা ও ইংরেজি ভাষার বিভিন্ন ভাষিক স্তর যেমন ধ্বনিসংগঠন, রূপ সংগঠন, বাক্য সংগঠনের সাদৃশ্য ও বৈসাদৃশ্য পর্যালোচনা করে এবং বৈসাদৃশ্যগুলিকে যথাসম্ভব কমিয়ে উভয়ের ভাষাতাত্ত্বিক কাঠামোকে সমরূপতায় উন্নীত করা, যার ফলে উভয় ভাষারই সমধর্মী ভাষাতাত্ত্বিক সূত্রাবলী তৈরি হবে এবং এগুলি পরিণতিকে গাণিতিক নিয়ম ও যুক্তি-সূত্রের অনুগামী হয়ে কম্পিউটার-বোধ্য হবে। এক্ষেত্রে অভিজ্ঞ অনুবাদক বাংলা থেকে ইংরেজি ভাষার প্রচলিত অনুবাদ যেসব নিয়মকানুন অনুসরণ করেন সেগুলিও সূত্রাবদ্ধ করা যেতে

পারে, যা চূড়ান্ত আলগোরিদম তৈরির ক্ষেত্রে সহায়ক হবে। তবে যে কোন ভাষার তিনটি ভাষিক স্তর যেমন ধ্বনি, রূপ ও বাক্য সংগঠনের ভাষা বৈজ্ঞানিক সূত্র উদ্ভাবন সম্ভব হলেও শব্দের অর্থের আলগোরিদম তৈরি এক জটিল ও প্রায় অসাধ্য কর্ম। কারণ শব্দের অর্থ ভাষার অন্য তিনটি স্তরের মত ততটা ইন্দ্রিয়গ্রাহ্য নয়, ফলে এর সূত্রায়ন ও সম্ভব নয়। তাছাড়া ভাষার অর্থ হচ্ছে সবচেয়ে অস্থির এলাকা। শব্দের অর্থ যুগে যুগে পাল্টে যায়, নতুন অর্থ ধারণ করে, কোন কোন শব্দের একাধিক অর্থ থাকে, আবার বিষয় ও পরিস্থিতি ভেদে একই শব্দ কখনও ভিন্ন অর্থ গ্রহণ করে। এসমস্ত কারণে বিশিষ্ট অর্থবিজ্ঞানী জিওফ্রে লিচ তাঁর semantics [১৯৭৪ : ১১-২৭] গ্রন্থে শব্দের সাত ধরনের অর্থের কথা বলেছেন, যা শব্দের ভিন্ন ভিন্ন পরিবেশে ভিন্ন ভিন্ন অবস্থানের কারণে ঘটে থাকে।

শব্দের অর্থের এই যে অনিত্য ধর্ম, তা যান্ত্রিক অনুবাদের ক্ষেত্রে বিজ্ঞানীদের রীতিমত হতোদ্যম করেছে এবং ১৯৫৫ সালের পর থেকে পৃথিবীতে যতগুলো যান্ত্রিক অনুবাদ কার্যক্রম পরিচালিত হয়েছে তার সিংহভাগের অসফলতা মূলত শব্দের সঙ্গে অর্থের সমন্বয়হীনতা এবং অর্থের আলগোরিদম প্রণয়নের ব্যর্থতার কারণে। তাই সংশ্লিষ্ট শব্দের সঠিক অর্থ নির্ধারণই হচ্ছে তখন যান্ত্রিক অনুবাদের ক্ষেত্রে বড় সমস্যা। এ সমস্যা সমাধানকল্পে যান্ত্রিক অনুবাদ বিজ্ঞানীরা নিরলস শ্রম সহযোগে সঠিক পন্থা উদ্ভাবনের জন্য নানাবিধ প্রচেষ্টা চালিয়ে যাচ্ছেন। এক্ষেত্রে তাদের প্রধান কাজ হচ্ছে দুই ভাষার অর্থানুর অভেদত্ব খুঁজে বের করার উপায় উদ্ভাবন করা।<sup>১৫</sup> তবে কম্পিউটারের শব্দার্থ সম্পর্কে দ্ব্যর্থকতা নিরসনকল্পে এবং বিভিন্ন দ্রব্য, বস্তু, ঘটনা বা বিষয় সম্পর্কে কম্পিউটারকে অবহিত করার জন্য তার স্মৃতিকোষে সংরক্ষণ করা যেতে পারে অগণিত তথ্য, যা সংশ্লিষ্ট বিষয় বা ঘটনা সম্পর্কে তাকে অবহিত করতে 'পূর্ববর্তী জ্ঞান' হিসেবে কাজ করবে। কারণ অভিজ্ঞ অনুবাদক যখন অনুবাদকর্মে প্রবৃত্ত হন তখন সমাজ ও পরিপার্শ্ব থেকে তাঁর চিন্তা ও অভিজ্ঞতার সমন্বয়ে আহৃত ও তাঁর মানবীয় মস্তিষ্কের স্মৃতিকোষে তথ্যরূপে সঞ্চিত 'পূর্ববর্তী জ্ঞানের'<sup>১৬</sup> দ্বারাই সংশ্লিষ্ট অনুবাদকর্মটির উদ্দেশ্য, মূল টেক্সটের লেখকের উদ্দেশ্য, শব্দ ও বাক্যের অর্থ পরস্পরা, বাহ্যিক অর্থ, অন্তর্নিহিত অর্থ, সামগ্রিক পরিপ্রেক্ষিত উপলব্ধি করতে সক্ষম হন। কিন্তু এক্ষেত্রে কম্পিউটারের সীমাবদ্ধতা হল মানব মস্তিষ্ক যেমন কোন ঘটনা বা বিষয় উপলব্ধি করতে তার স্মৃতিকোষে রক্ষিত সংশ্লিষ্ট পূর্ববর্তী জ্ঞানকে স্বয়ংক্রিয় প্রক্রিয়ায় প্রয়োগ করতে পারে কম্পিউটার তা এখনো পারেনি। হাজার তথ্যের ভিড় থেকে প্রাসঙ্গিক তথ্যটিকে উদ্ধার করে আনার ক্ষমতা কম্পিউটারের আজও সেভাবে হয়নি।<sup>১৭</sup> আর কম্পিউটারের সেই ক্ষমতা অর্জনের জন্য কম্পিউটার বিজ্ঞানে শুরু হয়েছে 'কৃত্রিম বুদ্ধিমতা' (artificial intelligence) নামে এক প্রাণসর ধারার।

ভাষা ব্যবহারের ক্ষমতা মানুষের বুদ্ধিমত্তার নিদর্শক।<sup>১৮</sup> কম্পিউটারে মানবীয় ভাষার প্রয়োগ এবং কম্পিউটারকে মানবীয় ভাষার ব্যবহারে বা তার মর্মার্থ উপলব্ধিতে সমর্থ করতে হলে এতে মানবীয় বুদ্ধিমত্তার অনুরূপ ক্ষমতার প্রয়োগ ঘটতে হবে। তাই কম্পিউটার বিজ্ঞানে অধ্যয়ন করা হচ্ছে কৃত্রিম বুদ্ধিমত্তার। কৃত্রিম বুদ্ধিমত্তা হচ্ছে তথ্য প্রক্রিয়াকরণের একটি প্রাথমিক গবেষণা ক্ষেত্র। এর লক্ষ্য হচ্ছে কিভাবে কম্পিউটার বুদ্ধির প্রয়োগে মানুষের সমকক্ষ হতে পারে বা তাকে ছাড়িয়ে যেতে পারে সেটি অনুসন্ধান করা অথবা বুদ্ধিমত্তার প্রয়োগ নির্ভর সমস্যা সমাধানের চেষ্টা করা।<sup>১৯</sup> মূলত কৃত্রিম বুদ্ধিমত্তা প্রদর্শনকারী পরিকল্পনা বা কৌশলসমূহ ও তার উন্নয়নকে উপজীব্য করে। এই গবেষণাক্ষেত্রের অধ্যয়নে সম্পৃক্ত হয়েছে বুদ্ধিমত্তার মৌল-প্রকৃতি উপলব্ধি সংক্রান্ত বৈজ্ঞানিক গবেষণা এবং কম্পিউটারকে অধিক 'সুবেশক' (smarter) করে তোলা সংক্রান্ত প্রকৌশলগত এবং উন্নয়নের উপাদানসমূহ। আশা করা যায় কৃত্রিম বুদ্ধিমত্তার গভীরতর চর্চা ও দ্রুত প্রসারের মাধ্যমে কম্পিউটার মানবীয় ভাষার মর্মার্থ উপলব্ধিতে সক্ষম হবে এবং স্বয়ংক্রিয় প্রক্রিয়ায় তার স্মৃতিকোষে সঞ্চিত পূর্ববর্তী জ্ঞানের সাহায্যে সংশ্লিষ্ট ঘটনা বা বিষয় বুঝার সামর্থ্য অর্জন করবে।

যান্ত্রিক অনুবাদ কার্যক্রমে মানবীয় ভাষার বিভিন্ন স্তরের আলগোরিদম প্রণয়নের কাজটি যখন সম্পন্ন হবে, অনুবাদ প্রক্রিয়া তখনই শুরু করা যাবে। এ ক্ষেত্রে উৎস ভাষার আলগোরিদম এবং লক্ষ্যভাষার আলগোরিদমকে কম্পিউটারের স্মৃতিকোষে ধারণ করা হবে এবং তাদের আন্তঃসম্পর্ক স্থাপিত হবে কম্পিউটারের কৃত্রিম ও যান্ত্রিক মধ্যবর্তী ভাষার মাধ্যমে। তারপর উৎসভাষার মূল টেক্সট যখন কম্পিউটারে 'ইনপুট' করা হবে তখন লক্ষ্যভাষায় তা যান্ত্রিকভাবে অনূদিত হয়ে আসবে। তবে এ প্রক্রিয়াটি সম্পন্ন হবে একেবারে যান্ত্রিক প্রক্রিয়ায় এবং আঙ্গিক ও যুক্তিগ্রাহ্য গণনার সহায়তায়। সমগ্র যান্ত্রিক অনুবাদ প্রক্রিয়াটি মোট তিনটি ধাপে<sup>২০</sup> সম্পন্ন হয়। এগুলি হচ্ছে—

- ক. প্রথম ধাপ—বিশ্লেষণ : যন্ত্র-সহায়ক পদক্রমিক বিশ্লেষণ প্রক্রিয়ার (passing) সাহায্যে উৎসভাষার বিশ্লেষণ।
- খ. দ্বিতীয় ধাপ—প্রেরণ : উৎসভাষা থেকে লক্ষ্যভাষায় তথ্য প্রেরণ।
- গ. তৃতীয় ধাপ—সংশ্লেষণ : লক্ষ্যভাষায় ফলাফল অর্জন।

তবে এই তিনটি ধাপের মধ্যে 'পূর্ব-সম্পাদনা' (pre-editing) এবং 'উত্তর-সম্পাদনা' (post-editing) নামক দুটি পর্যায় ক্রিয়াশীল থাকে। 'পূর্ব-সম্পাদনা' পর্যায়ে কম্পিউটার উপযোগী ভার্সন (version) উৎপাদনের জন্য নিয়ন্ত্রিত পদক্রম ও শব্দাবলী প্রয়োগ করে উৎস ভাষার মূল টেক্সটের পুনর্লিখন

হয়। আর 'উত্তর-সম্পাদনা' পর্যায়ে যন্ত্র-উৎপাদিত লক্ষ্যভাষার অমার্জিত (row) তথ্যকে নির্ভুল টেক্সট (error-free text) পরিণত করার জন্য সম্পাদনা করা হয়। এই দুটি সম্পাদনা পর্যায়ই কম্পিউটারের মাধ্যমে নিষ্পন্ন হয়ে থাকে। যান্ত্রিক অনুবাদ কার্যক্রমে উপরের যে তিনটি ধাপের উল্লেখ করা হল এই তিনটি ধাপের ধারাবাহিকতার কখনো সামান্য হেরফের ঘটে থাকে। যেমন-অনেক সময় উৎসভাষা থেকে লক্ষ্যভাষায় সরাসরি অনুবাদ কার্যক্রম সংঘটিত হয়ে থাকে, আবার একটি উৎসভাষার মূল টেক্সটকে যদি কয়েকটি ভিন্ন লক্ষ্যভাষায় রূপান্তর করতে হয়, তখন প্রথমে উৎস ভাষার টেক্সটকে একটি নিরপেক্ষ মধ্যবর্তী ভাষায় রূপান্তর করে তারপর অধিকতর জটিল প্রক্রিয়ায় ভিন্ন ভিন্ন লক্ষ্য ভাষায় অনুবাদ করতে হয়।

এই শতাব্দীর ষাটের দশকেও যান্ত্রিক অনুবাদ প্রকৃত অর্থে একটি আকর্ষণীয় অথচ অবাস্তব<sup>২১</sup> ঘটনা ছিল। কিন্তু কালক্রমে যন্ত্রের উত্তরোত্তর আধুনিকায়ন এবং বিভিন্ন মানবীয় ভাষার ভাষিক স্তর সমূহের যন্ত্রবোধ্য আঙ্গিক ও যুক্তিগ্রাহ্য ভাষাবৈজ্ঞানিক সূত্র আবিষ্কার হওয়ার ফলে যান্ত্রিক অনুবাদ ক্রমশ সফল পরিণতির দিকে যাচ্ছে এবং নব্বই দশকের শেষে এসে সফলতার সন্ধিক্ষণে দাঁড়িয়েছে। বর্তমানে জার্মান, জাপান, হল্যান্ড, যুক্তরাজ্য, যুক্তরাষ্ট্রসহ পৃথিবীর বিভিন্ন দেশের অসংখ্য বিশ্ববিদ্যালয়ে সংশ্লিষ্ট যান্ত্রিক অনুবাদের যে সমস্ত উচ্চতর গবেষণাকর্ম পরিচালিত হচ্ছে, তা হয়ত আগামী শতকের শুরুতেই ফলদায়ক হতে শুরু করবে। কিন্তু বাংলা ভাষা এক্ষেত্রে অনেক পিছিয়ে আছে। বাংলাদেশে এখন পর্যন্ত যান্ত্রিক অনুবাদের কোন বড় ধরনের কার্যক্রম শুরু হয়নি। তবে দেশের বিভিন্ন বিশ্ববিদ্যালয়ের কম্পিউটার বিজ্ঞান বিভাগে যান্ত্রিক অনুবাদ সহায়ক কতিপয় গবেষণাকর্ম পরিচালিত হচ্ছে বলে জানা যায়।<sup>২২</sup> মূলত কম্পিউটেশনাল ভাষাবিজ্ঞান যেহেতু ভাষাবিজ্ঞান ও কম্পিউটার বিজ্ঞানের একটি আন্তঃশাস্ত্রীয় নির্মিত তাই এক্ষেত্রে ভাষাবিজ্ঞানী ও কম্পিউটার বিজ্ঞানীর সম্মিলিত প্রচেষ্টাই পরিপূর্ণ সফলতা নিয়ে আসতে পারে।

কম্পিউটেশনাল ভাষাবিজ্ঞানের দ্বিতীয় উল্লেখযোগ্য ক্ষেত্র হলো কম্পিউটারের সাহায্যে কোন মানবীয় ভাষার ভাষা-উপান্তের অথবা কোন লেখকের মূল রচনার বিভিন্ন উপাদানের পরিসংখ্যানগত মূল্যায়ন। এই মূল্যায়নের মধ্যমে ঐ লেখক বা ভাষার বিভিন্ন নিজস্ব বৈশিষ্ট্য সম্পর্কে সম্যক পরিচয় লাভ করা যায়। কম্পিউটেশনাল ভাষাবিজ্ঞানের এই ধরনের গবেষণা কর্মে প্রচুর সংখ্যাভিত্তিক সূত্রের প্রয়োগ হয় বলে পরিসংখ্যানে পরিপূর্ণ জ্ঞান থাকা প্রয়োজন। এই গবেষণায় গবেষক সাধারণত মনবীয় ভাষা অথবা লেখকের রচনাকর্মকে বিশাল তথ্যভান্ডার বিবেচনা করে গবেষণা কর্মে অগ্রবর্তী হয়ে থাকেন।

মানবীয় ভাষার বিভিন্ন ভাষা-উপান্তের মধ্যে যে উপান্তটি নিয়ে কম্পিউটেশনাল ভাষাবিজ্ঞানে বেশী কাজ করা হয় সেটি হচ্ছে শব্দ। শব্দ যেহেতু ভাষার মূল বাহিকাশক্তি তাই কম্পিউটারে শব্দ বিষয়ক পরিসংখ্যানে ভাষার প্রাণধর্ম ধরা পড়ে। ধরা যাক, কম্পিউটারে বাংলা ভাষার শব্দের বিভিন্নধর্মী পৌনঃপুনিক বিশ্লেষণ করা হবে এবং এই ধরনের কাজের পর বাংলাভাষার শব্দের ভিন্ন ভিন্ন চারিত্র্য বৈশিষ্ট্য পাওয়া যাবে। যেমন- বাংলা ভাষায় প্রচলিত শব্দসমূহের মধ্যে তৎসম, তদ্ভব, বিদেশী, দেশী কোন ধরনের শব্দসংখ্যা কত তা জানা; বাংলা ভাষায় সবচেয়ে বেশী ব্যবহৃত পদ যেমন বিশেষ্য, বিশেষণ, ক্রিয়া, অব্যয়, সর্বনাম কি কি তা জানা; বাংলা ভাষায় ব্যবহৃত শব্দাবলীর মধ্যে কোন কোন বর্ণ অধিকবার আসে তা জানা; প্রাচীন, মধ্য ও আধুনিক এই তিন যুগে বাংলা ভাষায় কোন ধরনের শব্দ সবচেয়ে বেশী ব্যবহৃত হয়েছে তা জানা; ইত্যাদি আরও কত কি। কম্পিউটারে ভাষার শব্দ বিষয়ক গবেষণায় প্রাপ্ত তথ্য যেমন ভাষা শিক্ষার ক্ষেত্রে গুরুত্বপূর্ণ ভূমিকা রাখে তেমনি একটি ভাষার ব্যষ্টিক (micro) পরিকল্পনার ক্ষেত্রেও তাৎপর্যবাহী। তবে এ ধরনের গবেষণা বর্তমানে বর্ণনামূলক ভাষাবিজ্ঞানে ব্যাপকভাবে প্রয়োগ হচ্ছে।<sup>২৩</sup>

কোন লেখকের রচনাকর্মে ব্যবহৃত বিভিন্ন উপাদান যেমন শব্দের পৌনঃপুনিক বিশ্লেষণ ও কম্পিউটেশনাল ভাষাবিজ্ঞানের সাহায্যে করা যেতে পারে। এই ধরনের গবেষণার মাধ্যমে মূলত লেখকের রচনাকর্মের চারিত্র্যবৈশিষ্ট্যকে শনাক্ত করা যায়। পৃথিবীর বিভিন্ন ভাষার স্বনামখ্যাত লেখকরা কে কতটি ভিন্নধর্মী স্বতন্ত্র শব্দ (বিশেষ্য) ব্যবহার করে সরস রচনাকর্ম উপহার দিয়েছেন, কীর রচনাকর্মে বিশেষণ পদ বেশী, কে অব্যয়পদের প্রতি আসক্ত ইত্যাদি বিচিত্র গবেষণা আজ কম্পিউটেশনাল ভাষাবিজ্ঞানের আওতাধীন। যেমন ভক্তিপ্রসাদ মল্লিক (ফাল্গুন ১৪০২ : ৬০) সম্প্রতি কম্পিউটারে রবীন্দ্রনাথ ঠাকুরের গীতাঞ্জলি কাব্যসহ বিভিন্ন কাব্যগ্রন্থ বিশ্লেষণ করেছেন এবং গীতাঞ্জলি কাব্যের নিম্নোক্ত নতুন বৈশিষ্ট্য উদ্ভাবন করেছেন—

“CV প্যাটার্নের অক্ষরগুলির প্রচুরমাত্রায় ব্যবহার গীতাঞ্জলি কাব্যে এই উদাস্ত সুর সৃষ্টি করতে সমর্থ হয়েছে। CV প্যাটার্নের প্রাচুর্য বোঝায় কাব্য মূলত গীতিময়।”

লেখকের রচনারীতি সম্পর্কে উপরিউক্ত বৈশিষ্ট্য প্রদান ব্যতিরেকেও লেখকের ব্যবহৃত শব্দাবলী বা বিষয়সমূহের বর্ণানুক্রমিক সূচি তৈরি যেমন কম্পিউটারের মাধ্যমে সম্ভব তেমনি নির্দিষ্ট গ্রন্থের শব্দভিত্তিক নির্ঘন্ট তৈরির একঘেষে ও ক্লাস্তিকর কাজটিও আজ কম্পিউটেশনাল ভাষাবিজ্ঞানের কল্যাণে সম্পাদিত হচ্ছে।

কম্পিউটেশনাল ভাষাবিজ্ঞান নির্দিষ্ট লেখকের রচনাকর্মের বিভিন্ন উপাদান বিশ্লেষণ করে তাঁর রচনারীতি বা স্টাইল সম্পর্কে স্বকীয় বৈশিষ্ট্যসূচক যেসব সিদ্ধান্ত প্রদান করে তাদের সাহায্যে বাংলা সাহিত্যের অনেক বিতর্কিত সমস্যার সমাধান করা যেতে পারে। যেমন— বাংলা সাহিত্যে মধ্যযুগে রচিত অনেক উল্লেখযোগ্য রচনা আছে যেগুলো পুঁথির মাধ্যমে প্রাপ্ত এবং ঐ রচনাগুলো এখন যে লেখকের নামে পরিচিত পুঁথিতে সেই নাম না থাকায় রচনাকর্মটি আসলে সেই লেখকের কিনা তা নিয়ে বিভিন্ন রকম সংশয় আছে। এছাড়া বাংলা সাহিত্যে মধ্যযুগে একই নামে বিভিন্ন লেখকের বিভিন্ন রচনাও আছে। একই নামের বিভিন্ন লেখক প্রকৃত অর্থে একজন কিনা, বা সঠিক কতজন অথবা রচনাসমূহের চারিত্র্যবৈশিষ্ট্য সম্বন্ধে কিনা তা নিয়ে প্রচুর বিতর্ক আছে। ‘চন্দীদাস সমস্যার’ কথা এখানে উল্লেখ করা যেতে পারে। এখন এই সমস্যাও কম্পিউটেশনাল ভাষাবিজ্ঞানের কল্যাণে সমাধান হতে পারে। এক্ষেত্রে একজন লেখকের রচনামূল্যে নিজে বৈশিষ্ট্য নির্দেশক শব্দ বা উক্তি উপস্থিত আছে কিনা অথবা সেগুলি কিভাবে কতবার ব্যবহৃত হয়েছে তা পরীক্ষা করার জন্য অনুরূপ শব্দকোষ কম্পিউটারে ‘ইনপুট’ করা যেতে পারে। তারপর কম্পিউটার পরীক্ষণীয় অধ্যয়গুলির শৈল্পিক বিশ্লেষণ বা ভাষাবৈজ্ঞানিক সংরক্ষণ করে দেখবে এগুলি ‘ইনপুটকৃত’ শব্দকোষের সঙ্গে মিলে যায় কিনা। এতদ্ব্যতীত কবিতার ক্ষেত্রে ছন্দ মিলানো, প্রতিশব্দ বাছাই, কবিতার কাঠামোয় বিভিন্ন রকমফের সৃষ্টি, প্রবন্ধের সারাংশ লেখা আজকাল ইত্যাদিও কম্পিউটেশনাল ভাষাবিজ্ঞান করে দিতে পারে।

‘কম্পিউটেশনাল অভিধানতত্ত্ব’ কম্পিউটেশনাল ভাষাবিজ্ঞানের একটি গুরুত্বপূর্ণ গবেষণা ক্ষেত্র। অভিধান ভাষা তথা শব্দের শুদ্ধতা নির্ণয়ের প্রধান মাধ্যম, তাই জ্ঞানচর্চার প্রথম পর্যায়েও অভিধান প্রচলিত ছিল। কিন্তু অভিধান প্রণয়নের কাজটি ছিল যেহেতু সময় ও শ্রমসাপেক্ষ এবং সূক্ষ্মতা, নির্ভুলতা ও শুদ্ধতার প্রতিরূপক তাই এটি ছিল রীতিমত ক্লান্তিকর ও একঘেয়ে। ফলে অভিধান প্রণয়নের জন্যে নিষ্ঠাবান ও নিবেদিত প্রাণ সংকলকের শূন্যতা বরাবরই ছিল। কিন্তু কম্পিউটেশনাল ভাষাবিজ্ঞান এখন অভিধান প্রণয়নেও কার্যকরী ভূমিকা রাখছে। দ্বিভাষিক অভিধানে যেহেতু যান্ত্রিক অনুবাদের মতো উৎস ভাষা ও লক্ষ্যভাষা থাকে তাই কম্পিউটেশনাল অভিধানতত্ত্বেও উৎস ও লক্ষ্যভাষার আলগোরিদম প্রণয়নপূর্বক কম্পিউটারের স্মৃতিকোষে সঞ্চয় করতে হয় এবং পরবর্তীতে যান্ত্রিক অনুবাদের মতো স্বয়ংক্রিয় প্রক্রিয়ায় দ্বিভাষিক অভিধান তৈরি হয়।

কম্পিউটেশনাল ভাষাবিজ্ঞানের উপরিউক্ত গুরুত্বপূর্ণ গবেষণাক্ষেত্র ছাড়াও মানবীয় ভাষার বাচন প্রক্রিয়াকরণ, ভাষিক আচরণ অনুকরণের জন্য ভাষা-মডেল

প্রক্রিয়াকরণ ইত্যাদি নতুন নতুন ক্ষেত্র ও শাস্ত্রের গবেষণা কর্মের আওতাধীন হচ্ছে। মূলত যান্ত্রিক অনুবাদ প্রক্রিয়া সু-সম্পাদনের জন্য কম্পিউটেশনাল ভাষাবিজ্ঞান যাত্রা শুরু করলেও আজ এই শাস্ত্র মানবীয় ভাষার বিভিন্ন স্তরের অবয়ব ও তথ্যের বৈজ্ঞানিক বিশ্লেষণের বিশুদ্ধ ফলিত বিজ্ঞান হিসেবে সমগ্রতাস্পর্শী এবং মানববিদ্যার চারিদিক্য অতিক্রম করে ক্রমশ প্রযুক্তি অভিমুখী। কিন্তু তার পরেও ভাষাবিজ্ঞানী, কম্পিউটার বিজ্ঞানী এবং যন্ত্রের সুখম সমন্বয়েই গড়ে উঠেছে কম্পিউটেশনাল ভাষাবিজ্ঞান।

কম্পিউটেশনাল ভাষাবিজ্ঞানে যান্ত্রিক নিষ্পাদন অধিক পরিলক্ষিত হলেও তা কোনক্রমেই মানুষের মস্তিষ্কপ্রসূত মননশীলতা বর্জিত নয়। এবং এতে বৈজ্ঞানিক পদ্ধতিতে মস্তিষ্কপ্রসূত সৃজনশীল প্রক্রিয়াসমূহকে স্বয়ংক্রিয় করা হয় মাত্র। যেহেতু মানুষের মস্তিষ্কের সৃজনশীলতা অপরিসীম এবং কম্পিউটার ঐ মস্তিষ্কপ্রসূত সৃজনশীলতারই অনবদ্য আবিষ্কার তাই মানুষের প্রজ্ঞা ও মনীষার উপরই কম্পিউটার নির্ভরশীল। আবার যেহেতু যান্ত্রিক উৎকর্ষের কারণে কম্পিউটার মানুষের চেয়ে লক্ষগুণ 'চৌকস' সে কারণে কম্পিউটারই একমাত্র মানুষের সৃজনশীলতাকে জটিল গবেষণায় উজ্জীবিত করতে পারে। সর্বোপরি মানুষের স্বতঃস্ফূর্ততা 'সৃজনশীলতা' আর কম্পিউটারের স্বতঃস্ফূর্ততা 'যান্ত্রিক নিষ্পাদন'— এই দুই বিষয়কেই সংশ্লেষিত করে কম্পিউটেশনাল ভাষাবিজ্ঞান কার্যকর যান্ত্রিক অনুবাদ প্রক্রিয়ার পাশাপাশি ভাষা তথা সাহিত্যের বৈজ্ঞানিক বিশ্লেষণে বিপ্লবাত্মক পরিবর্তন নিয়ে আসতে পারে।

#### তথ্যসূত্র :

১. মুনীর চৌধুরী (১৩৬৯) : “সাহিত্য, সংখ্যাতত্ত্ব ও ভাষাতত্ত্ব”, বাংলা একাডেমী পত্রিকা, মাঘ-চৈত্র ১৩৬৯, পৃ. ৬৬
২. Bussmann, Hadumod (1996) : Routledge Dictionary of Language and Linguistics, Routledge Reference, London and New York, P. 91
৩. Crystal, David (1987) : The Cambridge Encyclopedia of Language, Cambridge University Press, Cambridge, P. 350
৪. Crystal, David (1987) : Ibid. P. 350
৫. Bridge, P. & S. Harlow (1995) : In Introduction of Computational Linguistics, Oxford
৬. Bussmann, Hadumod (1996) : Ibid. P. 291

৭. Malmkjaer, Kirsten (1995) : The Linguistics Encyclopedia, Routledge; London and New York, Reprint 1996, P. 30
৮. Dix, Alan. Finlay, Janet. Abowd, Gregory & Beale, Russell (1993) : Human-Computer Interaction, Prentice Hall, New York London P. 104
৯. Dix, Alan, Finlay, Janet, Abowd, Gregory & Beale, Russell (1993) : Ibid. P. 103
১০. সোমপ্রকাশ বন্দ্যোপাধ্যায় (১৯৯৩) : বোধ, বুদ্ধি এবং কম্পিউটার, আনন্দ পাবলিশার্স প্রাইভেট লিমিটেড, কলিকাতা, পৃ. ১৯
১১. Bussmann, Hadumod (1996) : Ibid, P. 15
১২. Claus, Volker & Schwilk, Andreas (1992) : Encyclopedia of Information Technology, Ellis Horwood, New York London, P. 10
১৩. সোমপ্রকাশ বন্দ্যোপাধ্যায় (১৯৯৩) : প্রাগুক্ত, পৃ. ১৯
১৪. Bussmann, Hadumod (1996) : Ibid. P. 169
১৫. সুনীল সেনগুপ্ত (১৯৮৯) : রবীন্দ্রনাথের ভাষাচিন্তা ও আধুনিক ভাষাবিজ্ঞান, নয়প্রকাশ, কলিকাতা ৭০০০০৬, পৃ. ৬৬
১৬. Yehoshua Bar-Hillel (1962) : "The Future of Machine Translation", 'Freeing the Mind', The Times Literary Supplement 1962, P. 34
১৭. সোমপ্রকাশ বন্দ্যোপাধ্যায় (১৯৯৩) : প্রাগুক্ত, পৃ. ৫৮
১৮. সোমপ্রকাশ বন্দ্যোপাধ্যায় (১৯৯৩) : প্রাগুক্ত, পৃ. ৫২
১৯. Claus, Volker & Schwilk, Andreas (1992): Ibid. P. 17
২০. Bussmann, Hadumod (1996) : Ibid. P. 291
২১. Yehoshua Bar-Hillel (1962) : Ibid. P. 36
২২. মাসিক 'কম্পিউটার জগৎ' ৬ষ্ঠ বর্ষ : ১ম সংখ্যায় (মে ১৯৯৬) প্রকাশিত "বিশ্ববিদ্যালয় পর্যায়ে গবেষণা : কম্পিউটার ও বাংলা ভাষা" শীর্ষক নিবন্ধের লেখক মুহাম্মদ শামীসুজ্জামানের তথ্য অনুযায়ী বাংলাদেশের বিভিন্ন বিশ্ববিদ্যালয় সমূহের কম্পিউটার বিজ্ঞান বিভাগে এ ধরনের গবেষণায় মোট ২১জন গবেষক কর্মরত আছেন। বিশ্ববিদ্যালয় ভিত্তিক এ পরিসংখ্যানটি হচ্ছে নিম্নরূপ—  
বাংলা প্রকৌশল বিশ্ববিদ্যালয়— ৫ জন, ঢাকা বিশ্ববিদ্যালয়ে— ৫ জন, রাজশাহী বিশ্ববিদ্যালয়ে— ১১ জন।
২৩. মুনীর চৌধুরী (১৩৬৯) : প্রাগুক্ত, পৃ. ৭০