Vol. 39 | No. 2 | 1996

Check for updates

সাহিত্য পত্রিকা

journal.bangla.du.ac.bd

কমপ্যুটারে রবীন্দ্র-কাব্যভাষা বিশ্লেষণ

Volume	39
Issue	2
Year	1996
ISSN	0558-1583
eISSN	3006-886X
Author(s)	ভক্তিপ্রসাদ মল্লিক
Published online	February 1, 1996
DOI	10.62328/sp.v39i2.3
Link to article	https://doi.org/10.62328/sp.v39i2.3
Pages	55-61
Publisher	University of Dhaka
Copyright	সাহিত্য পত্রিকা
Designed and Developed by	Zobayer Abdullah

Check for updates

কমপ্যুটারে রবীন্দ্র-কাব্যভাষা বিশ্লেষণ

ভক্তিপ্রসাদ মল্লিক

সাহিত্যকে আধার করে ভাষাতাত্ত্বিক ও ব্যাকরণগত গবেষণার কাজ চলে আসছে দীর্ঘকাল ধরে প্রাচ্য ও পাশ্চাত্ত্যের দেশগুলিতে। গবেষণার পদ্ধতি সর্বকালে ও সর্বদেশে একপ্রকার ছিলো না, এখনও তা নেই। পদ্ধতির পরিবর্তনের সঙ্গে যোগ হয়েছে মানসিক ও মানবিক দিকের প্রভাব ও প্রসার। আজ যান্ত্রিক সভ্যতার যুগে যন্ত্রেরও অনুপ্রবেশ ঘটেছে সাহিত্যের মতো শিল্পসমৃদ্ধ সূক্ষ্ম জগতের গবেষণা-ক্ষেত্রে।

কমপ্যুটার নামক যন্ত্রটির আবির্ভাবে দেশে দেশে জ্ঞানচর্চা নানা পথে এগিয়ে যেতে থাকে। অবশ্য এসব কিছুই ঘটেছে উন্নত দেশগুলিতে। আমরা সূক্ষ্ম যন্ত্র ব্যবহারের ক্ষেত্রে এখনো বেশ পিছিয়ে রয়েছি। যেমন আমাদের দেশে বিশেষজ্ঞের অভাব নেই কিন্তু অভাব অর্থের এবং সেই সঙ্গে উৎসাহের। উৎসাহের অভাবের ফলে পাশ্চাত্ত্য দেশগুলি যত তাড়াতাড়ি সবকিছু ধরে ফেলছে, আমরা সেখানে পরাভব মানছি।

পাশ্চাত্ত্যের দেশগুলিতে থাকাকালীন একসময় লেখকের ব্যক্তিগত অভিজ্ঞতা হয়েছিল যে এরা কমপ্যুটারকে ব্যবহার করতে চায় কেবল সামরিক প্রয়োজনে, জাপানে দেখেছি, এর বিপরীত চিত্র। কারণ হচ্ছে, জাপানে সামরিক খাতে ব্যয় মোট ব্যয়ের মাত্র ২ শতাংশ। ফলে মানবিক বিদ্যার গবেষণায় কমপ্যুটার জাপানে বিশিষ্ট ভূমিকা নিতে পেরেছে।

তোকিও থাকাকালীন তোকিও ইউনিভার্সিটি অব ফরেন স্টাডিজ-এর অধীন ইনস্টিটিউট ফর দ্য স্টাডি অব ল্যাঙ্গোয়েজেস অ্যান্ড কালচারস্ অব্ এশিয়া অ্যান্ড আফ্রিকা-র কমপ্যুটার বিভাগে বিশাল কমপ্যুটার গবেষণাযন্ত্র আমাকে বিস্মিত করেছিল। মনে পড়ে না ইউরোপ বা আমেরিকায় কোথাও শুধু ভাষা ও সংস্কৃতির গবেষণার জন্য এত বিরাটকায় কমপ্যুটার বিভাগ দেখেছি কিনা। জাপানে সর্ববৃহৎ ভাষা-সংস্কৃতি গবেষণাকেন্দ্র এটিই। একদা (১৯৮৫) এই বিভাগটি আমার কাছে

খুবই লোভনীয় বোধ হয়েছিল ভাষাগবেষণার চিন্তাপূরণ করার আংশিক ক্ষেত্র রূপে। এই বিভাগটিকে ভাষা গবেষণার ক্ষেত্রে কিভাবে ব্যবহার করা যেতে পারে তা নিয়ে ভাবনাচিন্তা করতে লাগলাম। মনে হল, আমাদের সাংস্কৃতিক জীবনের ফুসফুস রবীন্দ্রনাথ ঠাকুর। রবীন্দ্র-রচনাবলীর বৃহদারণ্যের যদি কিছু অংশ কমপ্যুটারের সাহায্যে ধরে নিয়ে গবেষণাকার্যে লাগানো যায়, তবে কিছু সাংস্কৃতিক দায়িত্ব পালন করার সুযোগ পাওয়া যেতে পারে।

রবীন্দ্রভাষা গবেষণার একটি সুপরিকল্পিত পরিকল্পনা প্রস্তুত করলাম যাতে স্বল্পকালের মধ্যে কমপ্যুটার-মাধ্যমে কিছু গবেষণাতথ্য বার করে নেওয়া যায়। এই প্রকল্পটি প্রণয়ন করতে লেখকের অতীতের অন্য একটি প্রকল্প যথেষ্ট সাহায্য করে। একদা Phonemic and Morphemic frequencies in Indian Languages-এর অন্তর্গত বাংলাভাষা নিয়ে পরিকল্পিত গবেষণার দায়িত্ব এসে পড়ে লেখকের ওপর। সেই সূত্রে আলোচ্য গবেষণার কয়েকটি বিশেষ দিক বিবেচনা করে কমপ্যুটারের কাছ থেকে একান্ত নির্ভরশীল Programme যা ভাষা-গবেষণার সহায়ক হবে তা রবীন্দ্র-গবেষণা পরিকল্পনার তালিকাভুক্ত করলাম। যেমন : রূপতাত্ত্বিক বর্ণনা, মুখশব্দসহ বাক্য বা বাক্যাংশ [concordance বা KWIC (key words in Context)] ইত্যাদি। তবে Concordence সর্বযুগে প্রয়োজনীয় বহুকিছু তথ্যের যোগান দেবে গবেষকের চিন্তা-ভাবনাকে কেন্দ্র করে। যদিও আমাদের আলোচনার বিষয় রবীন্দ্র-কাব্যভাষা, তথাপি প্রসঙ্গত উল্লেখ করা যায় তোকিওতে গল্পগুচ্ছের তথ্যসংগ্রহে (input) শুরুতে মন দিলাম। কবির গল্প রচনার প্রারম্ভিক কাল থেকে ১২৯৯ সন পর্যন্ত আঠাশটি ছোটগল্প গণকযন্ত্রে তোলা হল। যদিও রবীন্দ্রনাথ ছোটগল্প রচনা করেছেন একশটির কাছাকাছি। আর সেখানে ছড়ানো রয়েছে মণি-মুক্তার ভাণ্ডার। সময়াভাবে গল্পগুচ্ছ সম্পূর্ণ করা গেল না। গল্পগুচেছর এক চতুর্থাংশ গ্রহণ করে মুখ ফেরাতে বাধ্য হলাম কাব্যগ্রন্থাবলীর বিশ্লেষণে। রবীন্দ্রনাথ ছিলেন মূলত কবি। প্রথমে অবশ্য গীতাঞ্জলির কাজ শুরু করি। এরপর কাব্যগ্রন্থের বর্গীকরণ কালানুসারে শুরু করা হল। যেমন, শৈশবসঙ্গীত, বনফুল, কবি-কাহিনী, সন্ধ্যাসঙ্গীত, প্রভাতসঙ্গীত, ছবি ও গান, ভানুসিংহ ঠাকুরের পদাবলী, কড়ি ও কোমল, মানসী, সোনার তরী, নদী, চিত্রা, চৈতালি, কণিকা, কথা ও কাহিনী, কল্পনা, ক্ষণিকা, নৈবেদ্য, স্মরণ, শিশু, উৎসর্গ, খেয়া, গীতিমাল্য, গীতালি, বলাকা, পলাতকা, শিশু ভোলানাথ, পূরবী, লেখন, মহুয়া, বনবাণী, পরিশেষ।

ইলকা (আই এল সি এ এ) থেকে কমপ্যুটারের সাহায্যে পঞ্চাশোর্ধ কাব্যগ্রন্থের মধ্যে বত্রিশখানি কাব্যগ্রন্থের তথ্য সংগৃহীত হয়েছে, যা ধরা আছে Magnetic tape অথবা Floppy disc-এ। উচ্চ কারিগরী বিদ্যাকে আধার করে

আমরা এগোতে লাগলাম। এই গবেষণা-প্রক্রিয়াটিকে রূপ দিতে অধ্যাপক ৎসিয়োসি নারার ভূমিকা ভাষায় প্রকাশ করা সম্ভব নয়। তোকিওতে এ জাতীয় গবেষণা কখনো সম্ভবপর হত না যদি না অধ্যাপক নারার সাহায্য থাকত। তোকিও থেকে এ পর্যন্ত South Asian series-এ যে ক'খানি গ্রন্থ প্রকাশিত হয়েছে তার তালিকা দেওয়া হল : সন্ধ্যাসঙ্গীত, প্রভাতসঙ্গীত, গীতাঞ্জলি, ভানুসিংহ ঠাকুরের পদাবলী, মানসী, কথা ও কাহিনী, বলাকা, সভ্যতার সংকট। আনন্দের কথা, সভ্যতার সংকটের ক্ষেত্রে ইণ্ডিয়ান স্ট্যাটিস্টিক্যাল ইনস্টিটিউটের তন্দ্রা রাও কলকাতায় কমপ্যুটার প্রয়োজনায় যথেষ্ট সাহায্য করেন। কলকাতার ভূমিকা সম্পর্কে আলোচনায় আমরা পরে আসছি।

এবার বলি তোকিওতে আমরা কমপ্যুটার-সাহায্যে কোন্ কোন্ তথ্য আহরণ করি।

(১) আলোচ্য পাঠ্যগ্রন্থ কমপ্যুটার রীতি অনুযায়ী লিপিবদ্ধ করা। ভাষাটি কিন্তু কমপ্যুটারের নয়, সেখানে আমাদের বাংলা হরফই উল্লিখিত হয়। যেমন :

B E A A G T J O I 5 0 2 আমার মাথা নত করে দাও হে তোমার চরণ ধূলার তলে।

(২) বর্ণানুক্রমিক শব্দতালিকার বর্গীকরণ পরিসংখ্যা, বিভক্তি, প্রত্যয় ইত্যাদি সহ। যেমন :

ক্রমিক সংখ্যা	পরিসংখ্যা	শব্দ
1	1	অকাজের
2	1	অকারণে
3	1	অকিঞ্চণ

(৩) শব্দগুলির অধোগমনের তালিকা। যেমন :

ক্রমিক সংখ্যা	পরিসংখ্যা	শব্দ
1	205	আমার
2	199	তোমার
3	158	না

(৪) পরিসংখ্যাসহ বর্ণের তালিকা। যেমন :

ক্রমিক সংখ্যা	পরিসংখ্যা	বর্ণ
1	242	অ
2	1029	আ

3	551	ই

(৫) বর্ণের অধোগমন তালিকা। যেমন :

ক্রমিক সংখ্যা	পরিসংখ্যা	বর্ণ
1	1798	ম
2	1637	ব
3	1602	ত

তাছাড়া যুক্তব্যঞ্জনবর্ণও উল্লিখিত হয়েছে।

(৬) কনকরডান্সের তালিকা।

তালিকায় মুখশব্দসহ কোড উল্লিখিত হয়েছে। যেমন :

GTJ05003 শুধু কেবল সুরে বাজে/ অকাজের এই প্রাণ।
GTJ10802 কেবল তুমি আমি/যাব অকারণে ভেসে কেবল ভেসে,/
GTJ09006 এলে নিঃশেষে তায়/কর অকিঞ্চন। না থাকে তার মান অপমান,/

কনকরডান্সসহ পাঠসংকলন যাবতীয় শব্দগুলির বিভিন্ন পরিবেশে অর্থ, ব্যাকরণগত চেহারা, স্টাইলিস্টিক্স ইত্যাদির বিশ্লেষণ সম্ভবপর করে তুলবে। এখানে যা কিছু উল্লেখ করা হল তা গীতাঞ্জলি কাব্যগ্রন্থ থেকে। এইভাবেই বাকি বত্রিশখানি কাব্যগ্রন্থ, ছোটগল্প, প্রবন্ধের ভাষাতাত্ত্বিক-সংখ্যাতাত্ত্বিক পরিধি নির্ণয় করা হয়েছে। এই কাজগুলি করা হয়েছে তোকিওর ইলকাতে অধ্যাপক নারা ও অধ্যাপক সাকামোতোর ঐকান্তিক সাহচর্যে। গ্রন্থগুলির সম্পাদনার দায়িত্ব নিয়েছেন লেখকসহ এঁরা দু'জন। প্রথম পর্যায়ের কাজের এখানেই সমাপ্তি ঘটানো হল।

দ্বিতীয় পর্যায়ের কাজ শুরু হল কলকাতার ইণ্ডিয়ান স্ট্যাটিস্টিক্যাল ইনস্টিটিউটে। এই সংস্থার অধ্যাপক নিখিলেশ ভট্টাচার্য তাঁদেরই একজন যাঁরা বিশ্লেষণের পরিচালনার দায়িত্ব গ্রহণ করেন।

যা যা বিশ্লেষণ করা হয়েছে তার অতি সংক্ষিপ্ত উল্লেখ এখানে করা হচ্ছে।

(১) গীতাঞ্জলির মুখশব্দে মোট ৩০৯১টি শব্দ পাওয়া যায়। শব্দগুলিকে এইভাবে তালিকাভুক্ত করা হয়েছে।

ক্রমিক সংখ্যা	পরিসংখ্যা	শব্দ	শব্দের শ্রেণী	সমাসের শ্রেণী (কবিতা সংখ্যা)	ক্রিয়াপদের শ্রেণী
22	1	অতি-অসহায়	সংস্কৃত	কর্মধারয় (123)	
23	1	অতি-ইচ্ছার	সংস্কৃত	কর্মধারয় (2)	

162 । আসলে বাংলা চলিত

গীতাঞ্জলিতে ব্যবহৃত ক্রিয়াপদগুলির পরিসংখ্যা :

কাব্যিক 438, সাধু-চলিত 1167, চলিত1268 এবং সাধু 186।

সমাসের ক্ষেত্রে আমরা সংস্কৃত ভাষার ক্রাচ ব্যবহার করিনি। গীতাঞ্জলিতে বাংলা সমাস উল্লিখিত হয়েছে দ্বন্দ্ব, তৎপুরুষ, বহুব্রীহি, কর্মধারয়। তাছাড়া প্রত্যয়, বিভক্তি, উপসর্গের ক্ষেত্রেও বাংলা ভাষা যে একটি স্বতন্ত্র বলিষ্ঠ ভাষা সেই মত গীতাঞ্জলির ব্যাকরণ প্রস্তুত করেছি। যেমন, 'বিধান'-এর 'বি' আমাদের মতে একটি উপসর্গ নয়, 'বিধান' একটি স্বয়ং সম্পূর্ণ শব্দ (Unit word)।

(২) মূলশব্দ তালিকা : গীতাঞ্জলিতে মূলশব্দ ১২৩৮টি। সংস্কৃত (তৎসম) শব্দ ৩৫৬১, বাংলা (তদ্ভব) ৭৫৬১, মিশ্র ১৫০ ও বিদেশী ৪৯।

(৩) বিভিন্ন শ্রেণীর শব্দের শতকরা অনুপাত : সংস্কৃত শব্দ ৪৫·৩%, বাংলা শব্দ ৫৩·৩%, মিশ্র ১·৩%, বিদেশি শব্দ ০·১%, কবিতা সংখ্যা ৬৫-তে সংস্কৃত শব্দ পাই মাত্র ৭·৪%, বাংলা শব্দ ৯০·৫%, বিদেশী ২·১%, অর্থাৎ বাংলা শব্দের সংখ্যাধিক্য এ গানে খুবই বেশি। তেমনি কবিতা সংখ্যা ১৪৮-এ সংস্কৃত শব্দ ৭১·৯%, বাংলা শব্দ ২৬·৬%। ১৫৭টি গানে আমরা পেয়েছি মোট শতকরা হিসেব অনুসারে সংস্কৃত ৩১·৫%, বাংলা ৬৬·৮%, মিশ্র ১·৩%, আর বিদেশী ০·৪%। এই শব্দশ্রেণী বন্টন হয়ত উত্তরকালে ভাষাতাত্ত্বিক নানা গবেষণায় সাহায্য করতে পারবে।

(৪) বিভিন্ন শব্দের পরিসংখ্যা সম্পর্কে Zipf Law-এর উপযুক্ততার বিচার : এই আলোচনাটি সম্পূর্ণ রাশিবিজ্ঞান পদ্ধতিতে করা হয়েছে। যাঁরা বইটি দেখবেন তাঁরা এ সম্পর্কে ধ্যান ধারণা করতে পারবেন।

(৫) অক্ষর গঠনের পদ্ধতি ও তার পরিসংখ্যা :

V	যেমন এ, ও
VC	এক, এর ইত্যাদি
CVC	কাজ, জল ইত্যাদি
CV.	না, তো ইত্যাদি
CV.V	যায়, কায় ইত্যাদি
VV	আয়
CC.VC	প্রাণ, ত্রাণ ইত্যাদি
CCVV	প্রয়োজনে (CCV\V\V\CV\CV)

CCV ত্রাসে (CCV\CV) ত্রি(ভুবন)ঈত্যাদি

CV প্যাটার্নের অক্ষরগুলির প্রচুরমাত্রায় ব্যবহার গীতাঞ্জলি কাব্যে এই উদাত্ত সুর সৃষ্টি করতে সমর্থ হয়েছে। CV প্যাটার্নের প্রাচুর্য বোঝায় কাব্য মূলত গীতিময়। উল্লিখিত অক্ষর একাব্য ছাড়াও মান্য চলিত ভাষায় অবশ্য পাওয়া যায়। সুতরাং এখানে আমরা গীতিধর্মী কাব্যে ব্যবহৃত অক্ষরের চরিত্রগঠন পদ্ধতির কিছু পরিচয় পেলাম

(৬) বিভিন্ন কবিতায় ব্যবহৃত অন্ত্যমিল, অন্তর্মিল, ধ্রুবপদ ও অন্ত্য অসমমিল আলোচনায় এনেছি। অন্তর্মিলের ক্ষেত্রে যেমন 'তবু যা ভাঙাচোড়া ঘরেতে আছে সোরা।' 'আমি যে প্রাণ ভরি তাদের ঘৃণা করি।'

(৭) বিভিন্ন কবিতায় ব্যবহৃত ছন্দ।

মাত্রাবৃত্ত ছন্দ ৪১টি কবিতায়, স্বরবৃত্ত ১১২টিতে, অক্ষরবৃত্ত ৩টিতে, মাত্রাবৃত্ত অথবা অক্ষর ১টি কবিতায় পাই। মোট ১৫৭টি কবিতা ও গানের সংখ্যা।

(৮) শব্দদ্বৈত, শব্দত্রিত্ব, বিপরীতার্থক ও সমার্থক শব্দগুচ্ছসহ ধন্যাত্মক শব্দ যেমন 'গুরু গুরু', 'ছলছল', 'ঝরঝর' পাওয়া যায়।

(৯) বিভিন্ন কবিতায় ব্যবহৃত বিশেষণ-বিশেষ্য পদ। যেমন, 'অকূল তিমিরে,' 'অনাহত আঘাতে' 'অভ্রভেদী রথে' ইত্যাদি। এগুলি নানাভাবে বিভক্ত। যেমন, সংখ্যাবাচক বিশেষণ-বিশেষ্য, সর্বনামীয় বিশেষণ-বিশেষ্য ইত্যাদি।

(১০) সম্বন্ধ বিভক্তিযুক্ত আলঙ্কারিক বিশেষিত বিশেষ্য পদ। যেমন, 'আঁখির ক্ষুধা', 'মেঘের ভেলা', 'প্রাণের প্রদীপ', মানবজন্মতরীর মাঝি','সীমার বাঁধন' ইত্যাদি। প্রান্তিক পদগুলি বিশেষ্যবোধক, যদিও কয়েকটি পদ আভিধানিক অর্থে বিশেষ্য নয়।

(১১) ক্রিয়াপদগুলিরও বিশ্লেষণ করা হয়েছে : সাধারণ ক্রিয়া, যৌগিক ক্রিয়া, সংযোগমূলক ক্রিয়া এবং সংযোগমূলক যৌগিক ক্রিয়া।

(১২) বাক্যগঠনরীতি আমরা কেবল কর্তা, কর্ম ও ক্রিয়াপদের অবস্থান বিন্যাসের আলোচনা করেছি। গীতাঞ্জলিতে যতিচিহ্নিত বাক্যের সংখ্যা ৮৯৩টি। তার মধ্যে বাক্যাংশের গঠনপ্রকৃতি ৪৬৮ প্রকার এবং তার পরিসংখ্যা হচ্ছে ২১১৭। এর মধ্যে বাক্যাংশের যে 'প্রকার' (Type) অন্তত দু'বার পর্যন্ত পাওয়া গিয়েছে, সেই 'প্রকার'ই আমরা উল্লেখ করেছি, তার সংখ্যা ১১৯। তাছাড়া আরও ৩৪৯টি প্রকারের বাক্যাংশ পাওয়া গিয়েছে, যেগুলি মাত্র একবারই এসেছে। এতে করে প্রমাণ করে কবির অফুরন্ত কাব্যিক প্রতিভা। অর্থাৎ মাত্র ৮৯৩টি বাক্যে এত

বা ভিন্ন 'প্রকারের' বাক্যগঠনরীতি নিয়ে জাদুকরের মত খেলে গেছেন। যা অন্তত বাংলা ভাষার অন্য কোন কবির কাব্যে পাওয়া যায়নি।

এইভাবে কাব্যগ্রন্থগুলিকে নিয়ে বিশ্লেষণের মডেল তৈরি করা হয়েছে। যাবতীয় মডেল তৈরির পদ্ধতি এখানে উল্লিখিত হল না।

আমাদের বর্তমান পরিকল্পনা হচ্ছে রবীন্দ্রকাব্যগ্রন্থের Concordance তৈরি করা। তাছাড়া কালানুক্রমিক গবেষণার সুবিধার জন্য প্রতিটি কাব্যগ্রন্থের উপরে উল্লিখিত কমপ্যুটারের সাহায্যে Concordance তৈরি করা। অর্থাৎ কবির সাহিত্য জীবনের সৃষ্টিকে আমরা কয়েকটি খণ্ডে বিভক্ত করে আলোচনার সুযোগ নিতে পারি। রবীন্দ্রকাব্যগ্রন্থের গবেষণা কেবল রবীন্দ্রভাষাপরিক্রমার আধার হবে না। তা হবে বাংলাভাষার গবেষণার আধারও। বক্তব্য অধিক দীর্ঘ না করে এখানেই শেষ করছি। তবে বলে রাখা ভাল, ভবিষ্যৎ প্রজন্মের হাতে এর মূল্যায়নের দায়িত্ব ছেড়ে দিতে হবে। কোনো কোনো বিদ্বজ্জনের মতে সাহিত্যবিচারে ভাবালু অতিকথনের কাল শেষ হওয়ার সূচনা হল 'গীতাঞ্জলি'-র ভাষাতাত্ত্বিক-সংখ্যাতাত্ত্বিক গবেষণাকে কেন্দ্র করে।