সাহিত্য পত্রিকা

journal.bangla.du.ac.bd

Vol. 49 | No. 2 | 2012

Check for updates

অবয়ব ভাষাবিজ্ঞানের ভূমিকা : ঢাকাই ভাষা-অবয়ব বিশ্লেষণ

Volume	49
Issue	2
Year	2012
ISSN	0558-1583
eISSN	3006-886X
Author(s)	Syed Shahrier Rahman
Published online	February 1, 2012
DOI	10.62328/sp.v49i2.2
Link to article	https://doi.org/10.62328/sp.v49i2.2
Pages	২৯-৪১
Publisher	University of Dhaka
Copyright	সাহিত্য পত্রিকা
Designed and Developed by	Zobayer Abdullah

অবয়ব ভাষাবিজ্ঞানের ভূমিকা : ঢাকাই ভাষা-অবয়ব বিশ্লেষণ

Check for updates

সৈয়দ শাহরিয়ার রহমান*

অবয়ব ভাষাবিজ্ঞান (corpus linguistics)-এর মূল কাজ একটি ভাষা সম্পর্কিত সকল প্রকার পাঠকৃতির (text) বস্তুনিষ্ঠ ও নির্ভরযোগ্য নমুনা সংগ্রহ করে সেগুলো বিশ্লেষণের মধ্য দিয়ে ওই ভাষার প্রকৃতি ও পরিচয় নির্ণয় করা। অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণে নমুনা হিসেবে ভাষার কথ্য ও লেখ্য উভয় প্রকার পাঠকৃতিকেই বিবেচনা করা হয়ে থাকে। প্রাত্যহিক জীবনে ব্যবহৃত ভাষা-নমুনা, ভাষীদের পারস্পরিক কথোপকথন, আনুষ্ঠানিক কিংবা অনানুষ্ঠানিক সংলাপ, বক্তব্য, সাক্ষাৎকার– এগুলো যেমন ভাষা-অবয়বের অংশ তেমনি দৈনন্দিনের লিখিত নথি, পত্রিকা, সাহিত্য প্রভৃতিও এর আওতাভুক্ত। বিশেষ করে লোকসাহিত্যের সঙ্গে ভাষার স্বতঃস্ফূর্ত বিকাশের নিবিড় সম্পর্ক থাকায় বিভিন্ন লোকসাহিত্যিক অনুষঙ্গ অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণের উপকরণ রূপে বিবেচিত হয়ে থাকে। তবে, সাহিত্য ছাড়াও ভাষার বিভিন্ন কথ্য ও লেখ্য নমুনাকে বিশেষ গুরুত্বের সঙ্গে এই জাতীয় গবেষণায় পর্যবেক্ষণ করা হয়ে থাকে। বস্তুত অবয়ব ভাষাবিজ্ঞানের তাত্ত্বিক লক্ষ্য হিসেবে কাজ করে প্রধানত দুটি বিষয় : একটি হলো, ভাষীদের দৈনন্দিন যোগাযোগের প্রজ্ঞানমূলক (cognitive) বৈশিষ্ট্য কীরূপ তা বিচার করা এবং অপরটি হলো, ভাষা প্রক্রিয়াকরণে বিশেষত আধুনিককালে কৃত্রিম বুদ্ধিমত্তা সৃজনে সংশ্লিষ্ট ভাষা-অবয়ব (corpus)-কে ব্যবহার করা। এই সূত্র ধরে বলা যায় যে, অবয়ব ভাষাবিজ্ঞানের সঙ্গে তাই কম্পিউটার প্রযুক্তির ঘনিষ্ঠ সম্পর্ক বিদ্যমান। একদিকে, একটি ভাষার ভাষা-অবয়ব তৈরিতে এই প্রযুক্তি ব্যাপক সহায়তা প্রদান করে থাকে; অপরদিকে, প্রাকৃতিক ভাষা প্রক্রিয়াকরণে (natural language processing)-র প্রধান উপাত্ত উৎস হিসেবে সক্রিয় থাকে ওই ভাষার প্রস্তুতকৃত ভাষা-অবয়ব। এছাড়াও বিভিন্ন ধরনের ভাষা গবেষণা পরিচালনা করতে গেলে সংশ্লিষ্ট ভাষা-অবয়ব বিভিন্ন তাৎপর্যপূর্ণ তথ্য প্রদানে সহায়তা দান করে থাকে। সুতরাং, কোনো ভাষার একটি সাধারণ ভাষা-অবয়ব প্রস্তুত করা সম্ভব হলে, তা ওই ভাষার বিভিন্ন ভাষাবৈজ্ঞানিক বৈশিষ্ট্য নিরূপণে বিশেষ সাহায্য করে। উল্লেখ করা প্রয়োজন যে, এখন পর্যন্ত বাংলা ভাষায় বিশাল পরিসরে এ ধরনের কোনো উদ্যোগ গৃহীত হয়নি। যদিও উপনিবেশের কাল থেকেই বিভিন্ন মনীষীর ব্যক্তিগত দূরদর্শিতা ও উদ্যোগে বাংলা শব্দ সম্ভার তৈরির প্রচেষ্টা গৃহীত হয়েছে, তবুও এ কথা বলার অপেক্ষা রাখে না যে, দীর্ঘসময়ব্যাপী, সুপরিকল্পিত বিজ্ঞানসম্মত রাষ্ট্রীয় উদ্যোগ ব্যতীত কখনোই একটি ভাষার পূর্ণাঙ্গ ভাষা-অবয়ব প্রস্তুতি সম্ভব নয়। বিশেষ করে বাংলা ভাষার মতো একটি সমৃদ্ধ ভাষা, যার ভাষীর সংখ্যা বিশাল, সেরকম একটি ভাষার ভাষা-অবয়ব নির্মাণ নিঃসন্দেহে একটি বিরাট বিষয়। বর্তমান প্রবন্ধে আমরা মূলত অবয়ব ভাষাবিজ্ঞানের বৈশিষ্ট্য ও গুরুত্ব তুলে ধরে বাংলাদেশের ঢাকা নগরের অন্যতম উপভাষার অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণে সচেষ্ট

*সহকারী অধ্যাপক, ভাষাবিজ্ঞান বিভাগ, ঢাকা বিশ্ববিদ্যালয়।

হব। এ প্রবন্ধে একে 'ঢাকাই ভাষা' বলে অভিহিত করা হবে। বলে নেওয়া প্রয়োজন যে, ঢাকাই ভাষা বলতে মোটা দাগে মহানগরী ঢাকার 'পুরনো ঢাকা' অংশে ব্যবহৃত কথ্যভাষাকেই বোঝানো হয়ে থাকে। সাম্প্রতিককালে, বাংলাদেশ এশিয়াটিক সোসাইটির উদ্যোগে প্রকাশিত গ্রন্থে (ফিরোজা ইয়াসমীন (সম্পা.); ২০১০) উল্লিখিত ঢাকাই ভাষার সংগৃহীত বিভিন্ন ভাষা-নমুনা সংকলিত হয়েছে। যদিও একে ঢাকাই ভাষার কোনো পূর্ণাঙ্গ ভাষা-অবয়ব বলা যাবে না, তবে আমাদের জানা মতে, এটিই এখন পর্যন্ত এই উপভাষা বিষয়ক নমুনার সবচেয়ে বড় সংগ্রহ। বর্তমান প্রবন্ধে ঢাকাই ভাষার অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণের আকর রূপে ব্যবহৃত হবে উল্লিখিত গ্রন্থে উপস্থাপিত ও সংকলিত ভাষা-নমুনাসমূহ। আধুনিক দৃষ্টিকোণ থেকে অবয়ব ভাষাবিজ্ঞান বিষয়ক আলোচনা শেষে আমরা ঢাকাই উপভাষা বিশ্লেষণে নিয়েজিত হব।

ভাষা-অবয়ব : সংজ্ঞার্থ ও পরিধি

বর্তমান প্রবন্ধে 'ভাষা-অবয়ব' শব্দটি 'corpus' শব্দের পরিভাষা হিসেবে গৃহীত হয়েছে। লাতিন ভাষায় 'corpus' শব্দটির অর্থ 'দেহ-কাঠামো' বা 'চেহারা'। শব্দটির অর্থসংকোচন ঘটিয়ে এখন ভাষা-দেহের বিভিন্ন বৈশিষ্ট্য নিরূপণে 'corpus' কথাটি ব্যবহৃত হয়ে থাকে। এর সঙ্গেই সম্পর্কিত একটি পরিভাষা অবয়ব ভাষাবিজ্ঞান। তবে, একটি কথা মনে রাখা প্রয়োজন যে, ভাষাবিজ্ঞানের অন্যান্য শাখা, যেমন: সমাজভাষাবিজ্ঞান, মনোভাষাবিজ্ঞান প্রভৃতির সঙ্গে অবয়ব ভাষাবিজ্ঞানের একটি মৌলিক পার্থক্য হলো, এটি শাখা হিসেবে যতখানি বিশেষায়িত তার চেয়ে অনেক বেশি বিভিন্ন বিশেষায়িত শাখার উপাত্ত উৎস হিসেবে কাজ করে। অর্থাৎ, অবয়ব ভাষাবিজ্ঞানকে বলা যায় ভাষাবৈজ্ঞানিক গবেষণার একটি পদ্ধতিগত ভিত্তি (Leech; 1992 : 105)। ভাষাবিজ্ঞানের বিভিন্ন শাখাভুক্ত বিষয় নিয়ে গবেষণা করছে এমন গবেষকদের সহায়তা প্রদানের উপায় রূপেই মূলত ব্যবহৃত হয় ভাষা-অবয়ব এবং অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণ প্রক্রিয়ায় বিশ্লেষিত বিভিন্ন ভাষিক তথ্য। পূর্বের বক্তব্যের সূত্র ধরে বলা যায় যে, একটি ভাষা-অবয়ব হলো প্রকৃতপক্ষে পাঠকৃতির সমাবেশ [Meyer; 2004 : xii (preface)] যাদের ভাষাবৈজ্ঞানিক বিশ্লেষণে ব্যবহার করা হয়। আরও সুনির্দিষ্টভাবে বলা যায় যে, ভাষা-অবয়ব হলো একটি ভাষার সংগৃহীত পাঠকৃতিসমূহের অবয়ব, যাদের বৈশিষ্ট্য ও বিশেষত্ব নিরূপণই অবয়ব ভাষাবিজ্ঞানের দায়িত্ব। পাঠকৃতির বৈশিষ্ট্য নিরূপণের সূত্র ধরে আধুনিক কালে তাই অবয়ব ভাষাবিজ্ঞানের সঙ্গে কম্পিউটার প্রযুক্তির গড়ে উঠেছে এক অনিবার্য সাযুজ্য। কেননা, যে বিশাল পরিসরে উপাত্ত সংগ্রহের প্রত্যাশা সাধারণভাবে একটি ভাষা-অবয়বের থাকে, তার যথার্থ ও নির্ভুল উপস্থাপনে কম্পিউটারের ভূমিকা অপরিহার্য। এ প্রসঙ্গে স্মরণ করা যেতে পারে, মার্কিন ইংরেজির লিখিত রূপ থেকে সংগৃহীত দশ লক্ষ শব্দ নিয়ে গঠিত বিশ্বের প্রথম কম্পিউটার নির্দেশিত ভাষা-অবয়ব ব্রাউন করপাসের কথা। এতে গুরুত্ব দেওয়া হয়েছিল লিখিত মার্কিন ইংরেজির যত ধরনের ব্যবহার আছে সেগুলো সংগ্রহ করার প্রতি। এসব লিখিত রূপের প্রতিটির বৈশিষ্ট্য নিরূপণ এবং বিভিন্ন রূপের মধ্যে পারস্পরিক তুলনা করাই ছিল এর উদ্দেশ্য। তবে, ভাষা-অবয়ব হতে হলেই যে তাকে এতটা কাঠামোবদ্ধ

হতে হবে এমনটি বাধ্যতামূলক নয়। এর পরিচয় পাওয়া যায় আরেকটি ইংরেজি ভাষা-অবয়ব দ্য পেন ট্রিব্যাংক-এর সূত্র ধরে। এতে, ঊনপঞ্চাশ লক্ষ শব্দ রয়েছে যাতে সংগৃহীত হয়েছে শেয়ার বাজারের খবর, বেতার কথিকা, মার্ক টোয়েনের রচনাসহ আরও অনেক কিছু; আর এর সঙ্গে যুক্ত হয়েছে গোটা ব্রাউন করপাস। এই ভাষা-অবয়বটি আগেরটির মতো এতখানি কাঠামোবদ্ধ নয়, অনেক বেশি অসমবর্গীয় (heterogenous) উপাত্তে সমৃদ্ধ। সচেতনভাবেই এতে পাঠকৃতির সংখ্যার কিংবা ধরনের সৌষম্যের বিষয়টিকে বিবেচনায় নিয়ে আসা হয়নি (Marcus, Santorini, and Marcinkiewicz; 1993 : 314)। সুতরাং দেখা যাচ্ছে যে, উপাত্ত সংগ্রহের উদ্দেশ্য ও পরিকল্পনার ওপর নির্ভর করে ভাষা-অবয়বের গঠন পৃথক হতে পারে।

ভাষা-অবয়বের রয়েছে বিভিন্ন প্রকারভেদ। লেখ্য এবং কথ্য ভাষা-অবয়বের কথা ইতোমধ্যেই বলা হয়েছে। এছাড়াও উপাত্তের ধরন অনুসারে তৈরি করা যেতে পারে বিভিন্ন প্রকার ভাষা-অবয়ব। একটি ভাষার সকল ধরনের উপাত্তের সন্নিবেশ ঘটিয়ে তৈরি করা হয় সাধারণ ভাষা-অবয়ব (general corpus)। ব্রিটিশ ন্যাশনাল করপাসকে এই শ্রেণিভুক্ত করা যেতে পারে। কখনও আবার বিশেষ ধরনের ভাষা গবেষণায় ব্যবহৃত হয় বিশেষ ভাষা-অবয়ব (special corpus)। যেমন, ব্রায়ান জেমস ম্যাকহুইনির নেতৃত্বে শিশুদের ভাষা বিষয়ে গবেষণার জন্য প্রস্তুতকৃত একটি বহুল ব্যবহৃত ভাষা-অবয়ব হলো Child Language Data Exchange System তথা CHILDS (দ্র. MacWhinney 2000)। আবার উপভাষা চর্চার জন্য ব্যবহৃত হয় বিভিন্ন বিশেষায়িত ভাষা-অবয়ব। উদাহরণস্বরূপ বলা যায়, নরওয়েজিয়ান, সুইডিশ, ডেনিশ, ফারোএসে, আইসল্যান্ডিক, উভডালিয়ান ভাষার উপভাষাগুলো থেকে ২৭ লক্ষ শব্দ সংগ্রহ করে গঠিত এক বৃহৎ ভাষা-অবয়ব নরডিক ডায়লেক্ট করপাসের কথা। এছাড়াও প্রসঙ্গত উল্লেখ্য, জার্মানির ফ্রেইবুর্গ বিশ্ববিদ্যালয় উদ্যোগে ইংরেজির উপভাষাসমূহ নিয়ে তৈরি FRED নামক ভাষা-অবয়ব, জাপানের ৎসুকুবা বিশ্ববিদ্যালয়ের উদ্যোগে প্রস্তুতকৃত জাপানি কথ্য উপভাষার ভাষা-অবয়ব, এস্তোনিয়ার টার্টু বিশ্ববিদ্যালয়ের তত্ত্বাবধানে তৈরিকৃত করপাস অব এস্তোনিয়ান ডায়লেক্টস প্রভৃতি। উল্লিখিত বিশেষায়িত ভাষা-অবয়বের আওতায় আরও কয়েক প্রকার যুক্ত হতে পারে। এদের মধ্যে বিশেষভাবে উল্লেখযোগ্য, নমুনা ভাষা-অবয়ব (sample corpus); যার প্রধান উদ্দেশ্য হলো নির্দিষ্টসংখ্যক নমুনা সংগ্রহ করে সেগুলো নিবিড়ভাবে বিশ্লেষণ করা। নমুনা বিশ্লেষণ প্রাধান্য পাওয়ায় এতে নমুনার যোজন কিংবা বিয়োজন উভয়ই বিশেষভাবে সংরক্ষিত। অর্থাৎ, এ ধরনের ভাষা-অবয়ব একবার প্রস্তুত হয়ে গেলে এতে আর কোনো কিছু যোগ করা যায় না বা এ থেকে কোনো কিছু বাদ দেওয়া যায় না। জুরিখের ইংরেজি দৈনিকের ভাষা নিয়ে নির্মিত ভাষা-অবয়বটি এই শ্রেণির অন্তর্গত। কোনো একটি ভাষার সাহিত্যের নির্দিষ্ট প্রকার নিয়ে কালপর্বভিত্তিক ভাষা-অবয়ব তৈরি হলে একেও নমুনা ভাষা-অবয়ব রূপে চিহ্নিত করা যেতে পারে (যেমন: বাংলা ভাষার সাহিত্য নিয়ে হতে পারে 'উনিশ শতকের উপন্যাসের ভাষা-অবয়ব', 'বিশ শতকের ছোটগল্পের ভাষা-অবয়ব' প্রভৃতি)। একটি নমুনা ভাষা-অবয়ব যেমন গঠনগতভাবে অত্যন্ত রক্ষণশীল,

তেমনি এর বিপরীতে আরেক ধরনের ভাষা-অবয়ব দেখা যায় যেখানে প্রতিনিয়ত তথ্যভুক্তি সম্ভব। এ ধরনের ভাষা-অবয়বকে বলা যেতে পারে পর্যবেক্ষণকৃত ভাষা-অবয়ব (monitor corpus)। সময়ের সঙ্গে একটি ভাষার পরিবর্তনশীলতাকে পর্যবেক্ষণ এবং প্রয়োজনানুসারে দলিলীকরণ এ ধরনের ভাষা-অবয়বের লক্ষ্য। বিখ্যাত 'ব্যাঙ্ক অব ইংলিশ' এ জাতীয় ভাষা-অবয়ব। লক্ষণীয় যে, ভাষা-অবয়বের এই বিবিধ ধরন মূলত সুনির্দিষ্ট উদ্দেশ্যনির্ভর। একটি ভাষার বিভিন্ন বৈশিষ্ট্য অনুসন্ধানকে সামনে রেখে এক একটি ভাষা-অবয়বের নকশা প্রস্তুত করা হয়ে থাকে।

একটি ভাষার ভাষা-অবয়ব প্রস্তুতের ক্ষেত্রে এর অপরিহার্য বৈশিষ্ট্যসমূহের উপস্থিতি সুনিশ্চিত করা বিশেষভাবে প্রয়োজন। এই বৈশিষ্ট্যগুলোর মধ্যে রয়েছে এর সাংখ্যিক ব্যাপ্তি, গুণগত মান, প্রতিনিধিত্বশীলতা, তথ্যের সহজবোধ্যতা, নমুনা-সাম্য, তথ্যের উদ্ধার-সামর্থ্য (retrievability), যাচাইকরণ সুবিধা, ভাষার কালক্রমিক বিকাশধারার নিয়মিত পর্যবেক্ষণ-সামর্থ্য এবং দলিলীকরণ। বিষয়গুলো নিয়ে আলোচনা করা প্রয়োজন। বলা যেতে পারে, যে কোনো ভাষা-অবয়বের প্রথম লক্ষ্যই হলো সর্বোচ্চ সংখ্যায় নমুনা উপস্থাপন করা। অর্থাৎ, যে বিষয়গত সীমারেখা মেনে ভাষা-অবয়বটি তৈরি হচ্ছে তার আওতাভুক্ত প্রাসঙ্গিক সবকিছুকেই ওই ভাষা-অবয়বে যুক্ত করা। বলা যেতে পারে, প্রাসঙ্গিক ও কার্যকর উপাদানের সর্বোচ্চ সমাবেশ একটি ভাষা-অবয়বকে পূর্ণাঙ্গতর করে তুলতে সাহায্য করে। তবে, এই সাংখ্যিক ব্যাপ্তির ক্ষেত্রে মনে রাখা প্রয়োজন যে, কেবল প্রকৃত-নমুনাই একটি ভাষা-অবয়বে যুক্ত হতে পারে। লক্ষ রাখতে হবে, প্রতিটি নমুনা যেন স্বাভাবিক সংজ্ঞাপনের অংশ হয়। অর্থাৎ, কোনো বিশেষ বস্তু-পরিবেশে শর্ত-সাপেক্ষ কোনো কৃত্রিম ভাষা নমুনা যদি একটি সাধারণ ভাষা-অবয়বে যুক্ত হয় তবে তা ওই ভাষা-অবয়ব প্রস্তুতির মূল উদ্দেশ্যকে ব্যাহত করবে। এরই সূত্র ধরে গুরুত্বপূর্ণ হয়ে ওঠে নমুনার প্রতিনিধিত্বশীলতা গুণ। একটি ভাষা-অবয়বে প্রাসঙ্গিক নমুনার বৈচিত্র্য যত বেশি হবে, তত তা নির্ভরযোগ্য বিশ্লেষণকে নিশ্চিত করে তুলবে। এরই সঙ্গে প্রয়োজন সহজ ও সরলভাবে ভাষা নমুনার উপস্থাপন এবং বিভিন্ন নমুনার মধ্যে পারস্পরিক সমতা বজায় রাখা। কেননা, উপস্থাপিত নমুনা যদি সহজে বোধগম্য না হয় তবে তাকে নমুনা প্রক্রিয়াকরণে ব্যবহার করা কষ্টসাধ্য হয়ে পড়ে। একইভাবে, একটি ভাষা-অবয়বে যখন বিভিন্ন ধরনের নমুনার সমাবেশ বাঞ্ছনীয় হয়ে পড়ে, তখন প্রতিটি ধরনে সমপরিমাণ নমুনা সংগ্রহ বিশেষ জরুরি। তা না হলে, নমুনাসমূহের স্বাভাবিক বিন্যাস যেমন সম্ভব হয় না তেমনি ওই ভাষা-অবয়ব থেকে উপাদান সংগ্রহ করে বিশ্লেষণ করলে প্রায়শই সঠিক ফল আহরণ ব্যাহত হয়। এছাড়াও একটি ভাষা-অবয়বকে একজন ব্যবহারকারী যেন সহজে ব্যবহার করতে পারে, যে কোনো ধরনের ভাষাবৈজ্ঞানিক পরীক্ষণে কিংবা সংশ্লিষ্ট ভাষীদের কোনো বৈশিষ্ট্যকে যাচাই করে দেখবার জন্য ওই ভাষা-অবয়বের সাহায্য গ্রহণ যেন সম্ভব হয় তার জন্য প্রয়োজনীয় ব্যবস্থা গ্রহণ করা আবশ্যক। সেইসঙ্গে, সময়ের সঙ্গে তাল মিলিয়ে সংশ্লিষ্ট ভাষা-অবয়বটির পরিবর্তনশীলতাকে বিবেচনায় রেখে একে হালনাগাদ করা এবং

যথার্থভাবে দলিলীকরণের মাধ্যমে সর্বদা ব্যবহার উপযোগী রাখাও বিশেষভাবে প্রয়োজনীয়।

ভাষা-অবয়ব তৈরিতে কয়টি ভাষা অন্তর্ভুক্ত হবে তা-ও অত্যন্ত গুরুত্বপূর্ণ। এক, দুই এবং কখনও কখনও এরও বেশি ভাষা নিয়ে ভাষা-অবয়ব প্রস্তুত করবার প্রয়োজন দেখা দিতে পারে। সাধারণত, পরস্পর-সম্পর্কিত ভাষাগুলোকে একত্রে একই ভাষা-অবয়বের মধ্যে ধারণ করা হয়। তবে, কেবল একটি ভাষার বিবিধ বৈশিষ্ট্য নিয়ে গঠিত ভাষা-অবয়বের সংখ্যাই বিশ্বে প্রচুর। এক বা একাধিক যে সংখ্যক ভাষা নিয়েই ভাষা-অবয়ব তৈরি হোক না কেন, ওই ভাষা-অবয়বকে কার্যকর করে তুলতে প্রয়োজন ভাষাবৈজ্ঞানিক টীকাভাষ্যের (linguistic annotation)। বলা যেতে পারে, একটি সটীক ভাষা-অবয়ব (annotated corpus) যে কোনো ভাষাবৈজ্ঞানিক গবেষণায় অত্যন্ত গুরুত্বপূর্ণ ও তাৎপর্যপূর্ণ ভূমিকা পালন করতে সক্ষম। এর অর্থ এই নয় যে, প্রতিটি ভাষা-অবয়বকে বাধ্যতামূলকভাবে সটীক হতে হবে। তবে এ কথা অনস্বীকার্য যে, সটীক ভাষা-অবয়বের ভাষাবৈজ্ঞানিক গুরুত্ব অনেক বেশি। প্রসঙ্গত উল্লেখ্য যে, বর্তমান প্রবন্ধে আমাদের আলোচ্য ঢাকাই ভাষার আকর ভাষা-অবয়বটিও টীকাভাষ্যযুক্ত নয়। এই ভাষা-অবয়বটির টীকাভাষ্যের প্রকৃতি কীরূপ হতে পারে সে বিষয়ে আমরা পরবর্তী পর্যায়ে আলোকপাত করব।

ঢাকাই উপভাষা পরিচিতি

মূলত মিশ্র ভাষারূপ নিয়ে গড়ে উঠেছে ঢাকাই ভাষা। সমাজ-রাজনৈতিক পালাবদল ও প্রভাবের দীর্ঘ ঐতিহ্য লালনের মধ্য দিয়ে এই মিশ্রভাষা বাংলার একটি বিশিষ্ট উপভাষারূপে এখনও টিকে আছে। প্রসঙ্গত উল্লেখ করা প্রয়োজন যে, প্রাচীন নগর ঢাকার ভৌগোলিক সীমায় মূলত তিনটি ভাষারূপ লক্ষ করা যায়। এর একটি হলো অবিভক্ত বাংলার বিভিন্ন অঞ্চল থেকে পেশাগত, শিক্ষগত তথা জীবনমান উন্নয়নের উদ্দেশ্যে নগর-ঢাকায় আগত মানুষের নিজ নিজ ভাষার সংমিশ্রণে ও বৃহত্তর ঢাকা জেলার উপভাষাসমূহের প্রভাবে গড়ে ওঠা ভাষারূপ (এরই সূত্র ধরে কালক্রমে নতুন ঢাকায় ব্যবহৃত হতে শুরু করেছে একটি নাগরিক উপভাষা; যা আমাদের বর্তমান আলোচনার অন্তর্ভুক্ত নয়)। অপর দুটি রূপ হলো : "কুট্টি" ভাষা ও "সুখবাস" ভাষা। এই শেষের দুটি রূপ নিয়েই আলোচ্য ঢাকাই ভাষার সীমা নির্দেশিত। লক্ষণীয় যে, দুটি রূপেই রয়েছে বাংলা এবং উর্দুর সংমিশ্রণ। তবে, কুট্টিভাষায় রয়েছে বাংলার প্রাধান্য আর বিপরীতভাবে সুখবাসভাষায় রয়েছে উর্দুর প্রাধান্য (ফিরোজা ইয়াসমীন; ২০১০ : ১১৩)। দীর্ঘ কালপ্রবাহে এই ভাষায় বাংলাদেশের অন্যান্য উপভাষার বিচিত্র ভাষিক উপাদানের মিশ্রণ থাকারও সম্ভবনা রয়েছে। তাছাড়া নতুন ঢাকাকে ঘিরে ভিন্ন একটি উপভাষা কাঠামো গড়ে উঠায় এবং পুরনো ও নতুন ঢাকার মধ্যে প্রাত্যহিক ভাষা-সংযোগ (language contact) অব্যাহত থাকায় নতুন ঢাকার ভাষিক প্রভাবও আলোচ্য ঢাকাই ভাষায় থাকা অসম্ভব নয়। তবে, বর্তমান আলোচনায় আমরা এই উপভাষার সমাজভাষাবৈজ্ঞানিক ও তুলনামূলক - ঐতিহাসিক কোনো অন্তর্বিভাজনে (intra-division) নিয়োজিত না হয়ে একে একটি সমগ্রক হিসেবে গ্রহণ

করে আমাদের উপাত্ত উৎসে সংকলিত ভাষিক নমুনাসমূহের আলোকে এর অবয়ব-ভাষাবৈজ্ঞানিক বিবেচনায় সচেষ্ট হব।

ঢাকাই উপভাষার অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণ

একটি ভাষার অবয়ব ভাষাবৈজ্ঞনিক বিবেচনা মূলত অনেকগুলো কর্মধারার একটি সমন্বিত ফল। কর্মপন্থা নির্ধারণ, কৌশল নির্বাচন, নমুনা সংগ্রহ, প্রয়োজনানুসারে কম্পিউটার সহায়তার সঠিক ব্যবহার প্রভৃতির আশ্রয়ে এর পুরো কার্যক্রম পরিচালিত হয়। অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণের মূল কাজ হলো মাঠ পর্যায় থেকে সংগৃহীত উপাত্তকে সঠিক প্রক্রিয়ায় সুবিন্যস্ত করা এবং তা বিশ্লেষণ করা। ভাষা-অবয়ব প্রক্রিয়াকরণেরও বিভিন্ন পন্থা রয়েছে। এই বিবিধ পন্থাকে কার্যকরভাবে ব্যবহারের জন্য দুটি প্রধান উপায় (tool) গ্রহণ করা হয়ে থাকে; এগুলো হলো: সংখ্যাতাত্ত্বিক বিশ্লেষণ (statistical analysis) ও টীকাভাষ্য তৈরি (annotation)। এই দুটি উপায় ব্যবহার করেই ভাষা-অবয়ব সংশ্লিষ্ট বিভিন্ন প্রক্রিয়া কাজ করে থাকে। এই সকল প্রক্রিয়ার মধ্যে কয়েকটি উল্লেখযোগ্য হলো : শব্দ-দল তৈরি (word grouping), মুখ্যশব্দ সন্ধান (key-word search), সামঞ্জস্য বিধান (concordance), অভিধামূলক সন্নিধান (lexical collocation), রূপতাত্ত্বিক প্রক্রিয়াকরণ (morphological processing), ব্যাকরণিক সংবর্গের সূচিকরণ (parts-of-speech tagging), নমুনা বিন্যাসকরণ (parsing)। প্রসঙ্গত উল্লেখ্য যে, এই প্রক্রিয়াগুলোকে পূর্ণাঙ্গরূপে প্রয়োগ করবার সুযোগ আমরা বর্তমান আলোচনায় পাব না। কেননা, আমাদের ব্যবহৃত ভাষা-নমুনার আকার ছোট এবং উপাত্তের বৈচিত্র্যও ব্যাপক নয়। ফলে, নির্বাচিত কয়েকটি প্রক্রিয়া ব্যবহার করে আমরা ঢাকাই উপভাষার বিভিন্ন বৈশিষ্ট্য তুলে ধরবার চেষ্টা করব।

সংখ্যাতাত্ত্বিক বিশ্লেষণ

ঢাকাই ভাষার যে ভাষা-অবয়বটি আমরা এ প্রবন্ধে ব্যবহার করেছি, তার আকার ও ব্যাপ্তি বিষয়ে ইতোমধ্যেই উল্লেখ করা হয়েছে। এই বাস্তবতা মেনে নিয়েই আমরা ভাষা-অবয়বটির সংখ্যাতাত্ত্বিক বিবেচনায় অগ্রসর হব। মূলত, ভাষিক বিভিন্ন উপাদানের পরিসংখ্যান তুলে ধরে দেখতে চাওয়া হয় ওই উপাদানগুলো কীভাবে সংশ্লিষ্ট ভাষায় ব্যবহৃত হয়ে থাকে। এই জাতীয় পরিসংখ্যান তৈরির ক্ষেত্রে সবচেয়ে বেশি পর্যবেক্ষণ করা হয় ভাষায় কোনো একটি উপাদানের পৌনঃপুনিকতা কীরূপ। কেননা, একটি শব্দ, প্রত্যয় কিংবা কোনো ভাষিক উপকরণ ব্যবহারের শতকরা হার নির্ণয় করা সম্ভব হলে ওই ভাষা বিষয়ে বিভিন্ন তথ্য ও নির্দেশনা পাওয়া যেতে পারে।

বর্তমান আলোচনায় ব্যবহৃত ভাষা-অবয়বটি লক্ষ করলে দেখা যাবে যে, বিষয়বৈচিত্র্য অনুসারে এই তথ্য ভাণ্ডারটি চারটি অংশে বিভক্ত। এগুলো হলো : ১. প্রবাদ-প্রবচন, ২. কৌতুক, ৩. ছড়া, ৪. ভাষা দলিলীকরণ। এই অংশ চারটির মধ্যে প্রবাদ-প্রবচন, কৌতুক ও ছড়া অংশের প্রতিটি আবার দুই ভাগে বিভক্ত। এদের একটি হলো সরাসরি মাঠ পর্যায় থেকে সংগৃহীত নমুনা এবং অপরটি হলো গ্রন্থ থেকে সংগৃহীত নমুনা। উভয় প্রকার নমুনার

আলোকে নিচের সারণির মাধ্যমে এই ভাষা-অবয়বটির শব্দসংখ্যা সম্পর্কে ধারণা লাভ করা সম্ভব।

সারণি ১ : ভাষা-অবয়বে ব্যবহৃত শব্দ সংখ্যার শতকরা হার

নমুনা বৈচিত্র্য	মাঠ পর্যায় থেকে সংগৃহীত		গ্রন্থ থেকে সংগৃহীত	
	মোট শব্দ	শতকরা হার (%)	মোট শব্দ	শতকরা হার (%)
প্রবাদ-প্রবচন	১৭৮৮	১৪.১১	৫৫০	৩.০৪
কৌতুক	৫৩৬৫	৪২.৩৩	১৩৩০৯	৭৩.৬৪
ছড়া	১৩৫০	১০.৬৫	৪২১৩	২৩.৩১
ভাষা দলিলীকরণ	৪১৭০	৩২.৯০	-	-
মোট	১২৬৭৩		১৮০৭২	

লক্ষণীয় যে, মাঠ পর্যায়ে সংগৃহীত উপাত্তের তুলনায় গ্রন্থ থেকে সংগৃহীত উপাত্তের পরিমাণ বেশি। যদিও মাঠ পর্যায় থেকে অধিকতর উপাত্ত সংগৃহীত হলে তা সংশ্লিষ্ট ভাষার সাম্প্রতিকতম পরিস্থিতি পর্যবেক্ষণের ক্ষেত্রে অধিকতর সহায়ক হতে পারত। ওপরের সারণি থেকে আরও লক্ষ করা যায় যে, মাঠ পর্যায় এবং গ্রন্থ – উভয় উৎস থেকে সংকলিত উপাত্তসমূহের মধ্যে কৌতুক ও ভাষা দলিলীকরণ অংশ একত্রে মোট উপাত্তের সিংহভাগ বহন করছে। এটি এই ভাষা-অবয়বের অত্যন্ত ইতিবাচক একটি বৈশিষ্ট্য। কেননা, ব্যবহৃত ভাষা-অবয়বের প্রবাদ-প্রবচন এবং ছড়া অংশের তুলনায় কৌতুক এবং ভাষা দলিলীকরণ অংশে ভাষীরা সংগত কারণেই অপেক্ষকৃত বেশি সাবলীলতা প্রদর্শন করবেন বলে অনুমান করা যায়। আবার কৌতুকের তুলনায় ভাষা দলিলীকরণ অংশে তাদের আরও বেশি স্বতঃস্ফূর্ত ও স্বচ্ছন্দ থাকাই স্বাভাবিক। মনে রাখা প্রয়োজন যে, প্রথম তিনটি অংশ এদের গঠন বৈশিষ্ট্যের কারণেই শেষেরটির তুলনায় বেশি কাঠামোবদ্ধ। ফলে, ব্যবহারের সময় ভাষীরা অনিবার্যভাবেই ওই কাঠামোকে অনুসরণ করে থাকে। কিন্তু অপরপক্ষে ভাষা দলিলীকরণ অংশের তথ্য সংগ্রহের জন্য তথ্যদাতাদের নির্দিষ্ট গল্প বলবার জন্য অনুরোধ করা হয়েছে। একই সঙ্গে, পুরনো ঢাকার বিভিন্ন এলাকা থেকে তথ্যদাতা নির্বাচন করবার ফলে, এলাকাভিত্তিক ভাষা-বৈচিত্র্যের বিষয়টিও গুরুত্ব পেয়েছে। সর্বোপরি এই প্রক্রিয়ায় ভাষীদের স্বতঃস্ফূর্ত ভাষা ব্যবহারের নমুনা সংগ্রহ করা সম্ভব হয়েছে। এ কারণে, আমরা আমাদের আলোচনায় বিভিন্ন অবয়ব-ভাষাবৈজ্ঞানিক বিষয়ে আলোকপাতের ক্ষেত্রে কেবল ভাষা দলিলীকরণ অংশেরই সাহায্য নেব।

ভাষীদের ব্যবহৃত কথামালা (discourse) অনুসন্ধান করে শব্দ সংগ্রহের প্রচেষ্টা সুদীর্ঘকাল ধরে হয়ে এসেছে। হ্যানা ক্যাথরিন মুলেন্স কিংবা ঈশ্বরচন্দ্র বিদ্যাসাগরের সূত্র ধরে প্রমিত বাংলার ক্ষেত্রে এই মত যেমন প্রযোজ্য তেমনি ঢাকার উপভাষা বিষয়ক গবেষণার ক্ষেত্রেও এই মত প্রাসঙ্গিক। ঢাকাই ভাষার দলিলীকরণ অংশের শব্দ-দল

বিশ্লেষণের প্রসঙ্গে বলা যায় যে, উপভাষার শব্দদল তৈরির প্রচলিত প্রক্রিয়াটি (দ্র. রাজীব হুমায়ুন; ১৯৯৪, আবুল কালাম মনজুর মোরশেদ; ১৯৯৬, ফিরোজা ইয়াসমীন; ২০১০) এক্ষেত্রে অনুসরণ করা যেতে পারে। ঢাকাই ভাষার শব্দ তালিকাভিত্তিক আলোচনা যেহেতু ইতোমধ্যেই বিভিন্ন গবেষণায় (দ্র. মুহাম্মদ আবদুল হাই; বা.১৩৭২, রাজীব হুমায়ুন; ১৯৯৪ ও ২০১০, ফিরোজা ইয়াসমীন; ২০১০) বিশ্লেষিত হয়েছে সেহেতু আমরা এ বিষয়ে কোনো বিস্তৃত আলোচনায় যাব না। ভাষা দলিলীকরণ অংশ আরও বিশ্লেষণ করলে দেখা যায় যে, এতে বিভিন্ন বাক্যে ব্যক্তিবাচক সর্বনামের প্রয়োগ তাৎপর্যপূর্ণ ভূমিকা পালন করছে। বস্তুত এই ব্যক্তিবাচক সর্বনামগুলো গল্প তিনটির প্রধান চরিত্রগুলোর প্রতিনিধিত্ব করবার মধ্য দিয়ে তথ্যদাতাদের ভাষার অন্যতম মুখ্যশব্দ রূপে ব্যবহৃত হয়েছে। নিচের সারণিটি আমাদের এ বিষয়টি বুঝতে সহায়তা করবে :

সারণি ২ : ভাষা দলিলীকরণ অংশে ব্যক্তিবাচক সর্বনামের ব্যবহার

ব্যক্তিবাচক সর্বনাম	সংগ্রহ এলাকা				
	আলুবাজার	লালবাগ	সূত্রাপুর	গেণ্ডারিয়া	বংশাল
আমরা	২	০	২	২	৩
তুই	৩	০	০	০	০
তোমরা	০	০	০	১	১
তোমগো (তোমাদের)	০	০	০	১	০
তুমাগো (তোমাদের)	০	০	২	০	০
উই (সে)	৮	১	২১	১	০

লক্ষণীয় যে, এলাকা ভেদে সর্বনাম ব্যবহারের উল্লেখযোগ্য পার্থক্য বিদ্যমান। এক্ষেত্রে আমরা কেবল প্রথম পুরুষের এক বচন 'উই' বিষয়েই আলোকপাত করব। অন্য সর্বনামগুলো ব্যবহারের পৌনঃপুনিকতা স্বল্প এবং এলাকা ভেদে তারতম্যও বিশেষ তাৎপর্য বহন করছে না। কিন্তু 'উই'-এর ক্ষেত্রে দেখা যাচ্ছে যে, সূত্রাপুর এলাকায় এই সর্বনামের ব্যবহার সবচেয়ে বেশি। এর তুলনায় বংশাল এলাকায় 'উই'-এর ব্যবহার এর প্রায় এক-তৃতীয়াংশ এবং লালবাগ ও গেণ্ডারিয়া এলাকায় এর ব্যবহার প্রায় নগণ্য বললেই চলে। এক্ষেত্রে প্রশ্ন জাগে যে, ভাষা দলিলীকরণ অংশের ৪১৭০ শব্দের (দ্র. সারণি ১) মধ্যে এই স্বল্প পরিমাণ (সূত্রাপুরের প্রসঙ্গ বিবেচনা করেও) ব্যক্তিবাচক সর্বনাম ব্যবহারের কারণ কী? এর উত্তর সন্ধান করলে দেখা যাবে যে, নমুনা হিসেবে গৃহীত প্রায় সবগুলো গল্পেই তথ্যদাতারা সর্বনাম ব্যবহার অপেক্ষা মূল বিশেষ্যটি বারবার ব্যবহার করতেই অভ্যস্ত। এটি পুরনো ঢাকার অধিকাংশ এলাকার ঢাকাই ভাষা ব্যবহারকারী জনগোষ্ঠীর প্রবণতা কি না তা আরও ব্যাপক আকারে উপাত্ত সংগ্রহ করে অনুসন্ধান করা যেতে পারে। এই অনুসন্ধানের ফল যা-ই আসুক না কেন, উল্লিখিত আলোচনার সূত্র ধরে এ কথা নিঃসন্দেহে বলা যায় যে,

ব্যক্তিবাচক সর্বনামের ব্যবহার কিংবা তা ব্যবহার না করে মূল বিশেষ্য ব্যবহার – এ-সবকিছুই একটি ভাষা-অবয়বের বিশেষত্ব তথা মুখ্যশব্দের ব্যবহার ব্যাখ্যায় সহায়তা করে।

ভাষা-অবয়ব বিশ্লেষণে সামঞ্জস্য বিধানের ধারণাটি মূলত পাঠকৃতিতে ব্যবহৃত শব্দের পৌনঃপুনিকতার ভিত্তিতে নির্ধারিত হয়। এটি অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণের একটি অপরিহার্য অংশ, কেননা বিভিন্ন শব্দের পৌনঃপুনিকতা নিরূপণ সম্ভব হলে একটি ভাষায় শব্দ ব্যবহারেরর বিভিন্ন ধাঁচ (pattern) সম্পর্কে জানা সম্ভব হয়। বলা যেতে পারে, একটি শব্দ কোনো একটি পাঠকৃতির বিভিন্ন অংশের যেখানে যেখানে ব্যবহৃত হয়েছে সেখানে তাকে সুচিহ্নিত করে দেখা হয় যে সমগ্র পাঠকৃতিতে শব্দটির বিন্যাস সুষম ও সামঞ্জস্যপূর্ণ কি না। যদি তা না পাওয়া যায় তবে তারও কারণ অনুসন্ধান করা হয়। উদাহরণস্বরূপ, আমরা সারণি ৩ পর্যবেক্ষণ করতে পারি, যেখানে ব্যবহৃত সংযোজক শব্দের একটি তুলনামূলক চিত্র পাওয়া যায় :

সারণি ৩ : ভাষা দলিলীকরণ অংশে সংযোজক শব্দের ব্যবহার

সংযোজক শব্দ	সংগ্রহ এলাকা				
	আলুবাজার	লালবাগ	সূত্রাপুর	গেণ্ডারিয়া	বংশাল
মাগার	৭	১	১	১	১
হালা	১৩	২০	০	১	৬
আবে	১	০	০	০	১
অ্যালা	৩	০	০	০	০

লক্ষণীয়, 'হালা' শব্দটির পৌনঃপুনিকতা সবচেয়ে বেশি। কিন্তু এই আধিক্য আলুবাজার ও লালবাগ এলাকার ভাষায় যেভাবে দেখা যাচ্ছে তেমনি অন্যান্য এলাকার ভাষায় দেখতে পাওয়া যায় না। এর পিছনে দুটি কারণ থাকতে পারে : একটি হলো, ওই এলাকার লোকেরা তাদের বাচনে শব্দটি বেশি ব্যবহার করে থাকে; অথবা যাঁদের কাছ থেকে উপাত্ত সংগ্রহ করা হয়েছে তাদের নিজস্ব বাচনভঙ্গির অংশ হিসেবে এই শব্দটি বেশি ব্যবহৃত হয়। এ বিষয়ে সঠিক কোনো সিদ্ধান্তে উপনীত হওয়ার জন্যেও আরও বিশাল পরিসরে উপাত্ত সংগ্রহ করা প্রয়োজন।

টীকাভাষ্য তৈরি

যে ভাষার অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণ করতে হবে তার বিভিন্ন ভাষিক উপাদানের সটীক উপস্থাপন বিশেষভাবে জরুরি। এ প্রসঙ্গে ইতোমধ্যেই উল্লেখ করা হয়েছে যে, একটি ভাষা-অবয়ব কখনও কখনও টীকাভাষ্য ছাড়াও কেবল অবিন্যস্ত উপাত্তের সাহায্যেও গঠিত হতে পারে (যেমনঃ আমাদের বর্তমান আলোচনায় ব্যবহৃত ঢাকাই ভাষার ভাষা-অবয়বটি)। তবে, এ কথাও অস্বীকার করবার উপায় নেই যে কোনো অবয়ব-

ভাষাবৈজ্ঞানিক বিশ্লেষণে টীকাভাষ্যের প্রয়োজনীয়তা অপরিসীম। কেননা, এর সাহায্যেই পাঠকৃতিতে অবস্থিত বিভিন্ন ভাষিক উপাদানের প্রকৃতি সম্পর্কে অবগত হওয়া যায়। এরূপ টীকাভাষ্য নির্দেশের জন্য ব্যবহৃত প্রক্রিয়াগুলো ইতোমধ্যেই উল্লিখিত হয়েছে। নিচে নির্বাচিত কয়েকটি প্রক্রিয়া বিষয়ে আলোকপাত করা হলো।

ভাষা-অবয়ব বিশ্লেষণের অন্যতম প্রক্রিয়া হলো পাঠকৃতিতে সন্নিবেশিত শব্দসমূহের অভিধামূলক সন্নিধান। এর সাহায্যে একাধিক সন্নিহিত শব্দের পারস্পরিক সম্পর্ক অনুধাবন করা সম্ভব। পাঠকৃতিতে যে শব্দগুলো ব্যবহৃত হয়, এর কোনো কোনোটি তার পার্শ্ববর্তী শব্দের সঙ্গে গড়ে তোলে এক প্রকার সন্নিধানমূলক সম্পর্ক। ফলে, শব্দগুলো জোটবদ্ধভাবে নিজেদের বাগর্থকে প্রকাশ করে। উদাহরণস্বরূপ আমাদের আলোচ্য ভাষা দলিলীকরণ অংশে বহুল পরিমাণে ব্যবহৃত যৌগিক ক্রিয়াসমূহকে বিবেচনা করা যেতে পারে। 'করবার লাগলো', 'ফালাইতে লাগলো', 'বেড়াইতে লাগলো', 'তাকাইবার লাগলো', 'ঘুরবার লাগলো', 'ভাববার লাগলো' প্রভৃতি। 'লাগলো'-এর সঙ্গে উল্লিখিত উদাহরণগুলোর প্রতিটি এক ধরনের সন্নিধানমূলক সম্পর্ক তৈরি করেছে। অর্থাৎ, প্রতিটি শব্দজোটে 'লাগলো' থাকলেও প্রত্যেক ক্ষেত্রে এর অর্থ এক নয়। বলা যেতে পারে পাশের মূল ক্রিয়ার সঙ্গে সন্নিধানের ফলেই এটি একটি ক্রিয়াধারার (serial verb construction) সৃষ্টি করেছে এবং এটি অর্থপূর্ণ হয়ে উঠেছে। এভাবে, পাঠকৃতিতে ব্যবহৃত শব্দের অভিধামূলক সন্নিধানের মধ্য দিয়ে সংশ্লিষ্ট ভাষার অবয়ব সম্পর্কে জ্ঞান লাভ করা যেতে পারে।

পাঠকৃতিতে ব্যবহৃত শব্দ সম্পর্কে সকল তথ্য আহরণের জন্য আরও প্রয়োজন হয় এর রূপতাত্ত্বিক প্রক্রিয়াকরণের। এই প্রক্রিয়ার সাহায্যে বিভিন্ন ধরনের শব্দের বৈশিষ্ট্য সন্ধান করা হয়। এর মূল লক্ষ্য হলো প্রতিবেশ থেকে শব্দটিকে বিচ্ছিন্ন করে শব্দটির রূপতাত্ত্বিক গঠন, অর্থ, বাক্যতাত্ত্বিক ভূমিকা প্রভৃতি বিষয়ে আলোকপাত করা হয়। নিচে আমাদের আলোচ্য নমুনা উৎসের পাঠকৃতিতে ব্যবহৃত একটি শব্দ (ফিরোজা ইয়াসমীন; ২০১০ : ৪২০) বিশ্লেষণ করে দেখানো যেতে পারে :

বহির্রূপ	করতাছে
শব্দ-শ্রেণি	সমাপিকা ক্রিয়া
ধাতু	কর্
প্রত্যয় অংশ	- তাছে
বিধা নির্দেশক	- ত-
নিপাত অংশ	- আ- (ঘটমান)
ক্রিয়া সহায়ক	- ছ-
কাল নির্দেশক	- এ- (বর্তমান)
পুরুষ নির্দেশক	- ছে (প্রথম পুরুষ)
সম্মানসূচক চিহ্ন	নাই
বচন চিহ্ন	নাই (এক/বহু)
অর্থ	"সে/তারা করছে"

বস্তুত, পাঠকৃতিভুক্ত একটি শব্দের টীকাভাষ্য তৈরির একটি প্রধান পন্থা হলো এই রূপতাত্ত্বিক বৈশিষ্ট্য নিরূপণ করা। ওপরের নির্দেশিত কাঠামো ব্যবহার করে একটি রূপতাত্ত্বিক উপাদানের বিভিন্ন বৈশিষ্ট্য নির্ধারণ ও বিশ্লেষণ করা সম্ভব। তবে, শব্দসমূহের একক কাঠামো বিশ্লেষণের সাথে সাথে সামগ্রিক পাঠকৃতিতে এর অবস্থান চিহ্নিত করাও বিশেষভাবে প্রয়োজন। এ লক্ষ্যকে সামনে রেখে পাঠকৃতিভুক্ত প্রতিটি শব্দের ব্যাকরণিক সংবর্গ নির্ধারণ তাই জরুরি। বাংলা ভাষার ক্ষেত্রে আমরা নিচের শ্রেণিগুলোকে বিবেচনা করতে পারি :

বিশেষ্য [বি], সর্বনাম [সনা], বিশেষণ [বিণ], সমাপিকা ক্রিয়া [সকা ক্রি], ক্রিয়া বিশেষণ [ক্রি বিণ],
অসমাপিকা ক্রিয়া [অসকা ক্রি], পরপ্রত্যয় [প প্র], অব্যয় [অ], সংখ্যাবাচক পদ [স প], প্রশ্নবাচক পদ [প্রবা প]

উল্লিখিত এই শ্রেণিসমূহের আলোকে আমরা আমাদের আলোচ্য ভাষা দলিলীকরণ অংশের একটি নির্বাচিত অনুচ্ছেদের ব্যাকরণিক সংবর্গ সূচিকরণে সচেষ্ট হব। অনুচ্ছেদটি (ফিরোজা ইয়াসমীন; ২০১০ : ৪১৫) নিম্নরূপ :

[ক্রি বিণ]গরমকালে [ক্রি বিণ]একদিন [স প]একটা [বি]কাউয়ার [বণ]বহুত [বি]পানির [সকা ক্রি](পিয়াস লাগলো)। [বি]কাউয়াটা [বি]পানির [প প্র]লাইগ্যা [ক্রি বিণ](এদিক ওদিক) [সকা ক্রি](ছুটতে লাগলো)। [ক্রি বিণ](বহুতক্ষণ পরে) [বি]কাউয়ায় [সপ]একটা [বি]পানির [বিণ]মাইটা [বি]কলসি [সকা ক্রি](খুইজা পাইলো)। [বি]কাউয়াটা [প প্র]ওমনেই [সনা]ওই [বি]কলসিটার [প প্র]উপরে [সকা ক্রি](গিয়া বইলো)। [অ]মাগার [বি](কাউয়া হালায়) [সকা ক্রি]দেখলো [বি]কলসির [বি]পানি [বিণ]বহুত [বি]নীচে। [বি]মুখ [প প্র]দিয়া [অসকা ক্রি](লাগুর পাওয়া) [সকা ক্রি](যাইবো না)। [সনা]হালায় [ক্রি বিণ]ওখন [প্রবা প]কি [সকা ক্রি]করবো? [ক্রি বিণ](এদিক ওই দিক) [অসকা ক্রি](চাওয়া চাওয়ি কইরা) [ক্রি বিণ]একটু [বিণ]দূরে [স প]কয়েকটা [বি](পাথরের টুকরা) [সকা ক্রি](দেখতে পাইলো)। [বি](কাউয়া হালার) [বি]দেমাকে [বি]বুদ্ধি [সকা ক্রি]আইলো, [অ]যদি [বি](পাথরের টুকরাগুলা) [বি]কলসির [বি]পানির [প প্র]মধ্যে [সকা ক্রি](ফালান যায়), [অ]তাইলে [সনা]হালায় [বি]পানি [ক্রি বিণ]ওপরে [সকা ক্রি] (উইঠা আইবো) [অ]আর [অ]তখন [বি]কাউয়ায় [বি]পানি [সকা ক্রি](খাইতে পারবো)। [অ]যেমন [বি]ভাবনা, [অ]ওমনি [বি]কাম, [বি](কাউয়া হালায়) [স প]একটা [স প]একটা [অসকা ক্রি]কইরা [বি](পাথরের টুকরা) [বি]কলসির [প্র প]মধ্যে [সকা ক্রি](ফালাইতে লাগলো)। [বিণ]কিছু [বি]সময়ের [প্র প]মধ্যেই [বি]কলসির [ক্রি বিণ]নীচের [বি]পানি [ক্রি বিণ]উপরে [সকা ক্রি](উইঠা আইলো)। [বি](কাউয়া হালায়) [বি]মন [অসকা ক্রি]ভইরা [বি]পানি [অসকা ক্রি]খায়া [সকা ক্রি](উইড়া গেল)।

একই প্রক্রিয়া ব্যবহার করে সমগ্র ভাষা-অবয়বের শব্দসমূহকে তাদের ব্যাকরণিক সংবর্গ অনুসারে বিন্যস্ত করা যেতে পারে। এই কাজ সমাপ্ত হলে এর সূত্র ধরে বিভিন্ন পরিসংখ্যানমূলক তথ্যও পাওয়া সম্ভব। জানা সম্ভব, ওই ভাষা-অবয়বে বিভিন্ন ব্যাকরণিক সংবর্গের ব্যাবহারিক পৌনঃপুনিকতা কীরূপ। একই সঙ্গে, বিভিন্ন ব্যাকরণিক সংবর্গের ব্যবহার-মাত্রার তুলনা করে তাদের বেশি বা কম ব্যবহৃত হওয়ার কারণ অনুসন্ধান করা যেতে পারে।

ওপরে নির্দেশিত এই টীকাভাষ্যে ব্যবহৃত শব্দসমূহের বৈশিষ্ট্য সম্পর্কে অবগত হওয়া সম্ভব হলেও শব্দগুলোর সমন্বয়ে সৃষ্ট বাক্যে তাদের পদগত অবস্থান ও বিন্যাসের বিষয়টি

স্পষ্ট হয়ে ওঠে না। এ কারণেই প্রয়োজন হয় নমুনা বিন্যাসকরণের। বলা যেতে পারে, এই নমুনা বিন্যাসকরণের প্রধান কাজ হলো পাঠকৃতির পদবিন্যাস শনাক্ত করা (McEnery and Wilson; 1996:178)। এর সাহায্যে পাঠকৃতির রূপবাক্যিক (morphosyntactic) উপাদানগুলোকে সুনির্দিষ্ট পদ্ধতিতে বিন্যস্ত করা হয়। বাক্যতাত্ত্বিক নির্দেশনা অনুসরণ করেই এই বিন্যাস প্রক্রিয়া সম্পন্ন হয় বলে বিবিধ বাক্যতাত্ত্বিক পন্থা দ্বারা এটি প্রভাবিত। তবে, সাধারণভাবে বলা যায় যে, একটি বাক্যিক অধিক্ষেত্রের (domain) সদৃশ তথ্যসমূহকে একত্রে বন্ধনীভুক্ত করে উপস্থাপনের মধ্য দিয়েই এই বিন্যাস প্রক্রিয়াটি সাধিত হয়। নিচে আমাদের উপাত্ত উৎস থেকে একটি বাক্য নির্বাচন করে এর নমুনা বিন্যাসকরণ প্রক্রিয়াটি দেখানো যেতে পারে (ফিরোজা ইয়াসমীন; ২০১০ : ৪২০) :

[$_{বা}$[$_{বি\ প}$[এক বান্দরের]$_{বি\ প}$] [$_{বিণ\ প}$[বহুত]$_{ক্রি}$ [খিদা পাইছে] ।]$_{ক্রি\ প}$]$_{বা}$]

লক্ষণীয় যে, বিন্যাস প্রক্রিয়াটির মূলনীতি হিসেবে কাজ করেছে প্রতিবেশমুক্ত পদ সাংগঠনিক ব্যাকরণ (context free phrase structure grammar)। মূলত, ভাষার অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণে এই পদ্ধতিই সবচেয়ে বেশি ব্যবহৃত হয়ে থাকে। নমুনার বিন্যাসের এই প্রক্রিয়া একটি ভাষা-অবয়বের শব্দস্তর, পদস্তর ও বাক্যস্তরের বিভিন্ন তথ্য তুলে ধরতে সহায়তা করে। পাঠকৃতিতে বিন্যস্ত সকল বাক্যের বিন্যাসকরণ সম্ভব হলে তা সমগ্র ভাষা-অবয়ব সংশ্লিষ্ট বৈচিত্র্যপূর্ণ ও পরস্পরসম্পর্কিত বিভিন্ন তথ্য প্রদান করে।

উপসংহার

ওপরে আলোচিত বিভিন্ন প্রক্রিয়া ছাড়াও অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণের আরও নানা প্রক্রিয়া রয়েছে। বিশেষ করে কম্পিউটারের সহায়তা গ্রহণের মধ্য দিয়ে কীভাবে ভাষা-অবয়ব সংশ্লিষ্ট তথ্য আহরণ করা যায় তা নিয়ে ব্যাপক আলোচনার সুযোগ বিদ্যমান। কিন্তু এই প্রসঙ্গটির কারিগরি সূক্ষ্মতা ও ব্যাপ্তির কথা চিন্তা করেই এই আলোচনায় নিয়ে আসা হয়নি। বর্তমান প্রবন্ধে আমরা মূলত দেখাতে চেয়েছি একটি ভাষা-অবয়বের বিশ্লেষণে কোন দিকগুলো সবচেয়ে গুরুত্বসহকারে বিবেচিত হয়ে থাকে এবং এর সূত্র ধরে ঢাকাই ভাষার অবয়ব-ভাষাবৈজ্ঞানিক বিশ্লেষণের উল্লেখযোগ্য দিকসমূহ কী কী। তবে, এ ক্ষেত্রেও একটি সীমিত পরিসরই আমরা আমাদের আলোচনায় ব্যবহার করেছি। আমাদের ব্যবহৃত আকর গ্রন্থে সংকলিত প্রবাদ-প্রবচন, কৌতুক ও ছড়া বিষয়ে আমরা কোনো বিস্তৃত আলোচনায় যাইনি। কেননা, এগুলোর প্রত্যেকটি নিয়েই স্বতন্ত্র আলোচনার সুযোগ বিদ্যমান। পুনরুল্লেখ করা যেতে পারে যে, আকর গ্রন্থের ভাষা দলিলীকরণ অংশটি যেহেতু ঢাকাই ভাষার কথ্যরূপের একটি স্বতঃস্ফূর্ত সংগ্রহ সেহেতু একেই আমরা অবলম্বন করতে চেয়েছি। ভবিষ্যতে এই দলিলীকরণ অংশটি আরও বড় রূপ পেলে বিশেষত এতে আন্তর্জাতিক ধ্বনিলিপির সাহায্যে প্রতিবর্ণীকৃত উপাত্ত উপস্থাপিত হলে এই উপভাষার ধ্বনিতাত্ত্বিক বৈশিষ্ট্য নিরূপণসহ আরও বিভিন্ন বিষয়ে ব্যাপকতর গবেষণার সুযোগ তৈরি হবে। সেইসঙ্গে বলা যেতে পারে যে, ঢাকাই ভাষা সংগ্রহের উদ্যোগে উৎসাহী হয়ে বাংলা ভাষার বিভিন্ন উপভাষাসহ প্রমিত বাংলার জন্য যেমন ভাষা-অবয়ব প্রণয়নের প্রচেষ্টা গ্রহণ

করা প্রয়োজন, তেমনি এই সংশ্লিষ্ট উপাত্তসমূহকে অবয়ব-ভাষাবৈজ্ঞানিক দৃষ্টিকোণ থেকে আরও বিস্তৃত পরিসরে বিশ্লেষণ করা আবশ্যক। এই পরস্পর-সম্পর্কিত উদ্যোগসমূহ কার্যকরভাবে গ্রহণ করা সম্ভব হলে বাংলা ভাষার গবেষণা আরও সমৃদ্ধ হবে এবং এই ভাষা সম্পর্কিত বহু অনালোকিত দিক উন্মোচিত হবে।

আকরগ্রন্থ

ফিরোজা ইয়াসমীন (সম্পা.)। (২০১০)। *ঢাকাই উপভাষা : প্রবাদ প্রবচন কৌতুক ছড়া*, ঢাকা : বাংলাদেশ এশিয়াটিক সোসাইটি।

সহায়ক গ্রন্থ ও প্রবন্ধ

আবুল কালাম মনজুর মোরশেদ। (২০০৬)। "ঢাকাবাসীদের ভাষাবিন্যাস", *রবীন্দ্রনাথ, নজরুল ও অন্যান্য প্রসঙ্গ* গ্রন্থের অন্তর্গত, ঢাকা।

ফিরোজা ইয়াসমীন। (২০১০)। "পুরোনো ঢাকার মিশ্র উপভাষা", ফিরোজা ইয়াসমীন (সম্পা.) *ঢাকাই উপভাষা : প্রবাদ প্রবচন কৌতুক ছড়া* গ্রন্থের অন্তর্গত, ঢাকা : বাংলাদেশ এশিয়াটিক সোসাইটি।

মুহম্মদ আবদুল হাই। (বা.১৩৭২)। "ঢাকাই উপভাষা", *সাহিত্য পত্রিকা*, ৯ম বর্ষ, ১ম সংখ্যা।

রাজীব হুমায়ুন। (১৯৯৪)। "পুরোনো ঢাকার ভাষা", *নিবন্ধমালা*, অষ্টম খণ্ড, ঢাকা: ঢাকা বিশ্ববিদ্যালয়।

রাজীব হুমায়ুন। (২০১০)। "পুরোনো ঢাকার বাংলা উপভাষা", ফিরোজা ইয়াসমীন (সম্পা.) *ঢাকাই উপভাষা : প্রবাদ প্রবচন কৌতুক ছড়া* গ্রন্থের অন্তর্গত, ঢাকা : বাংলাদেশ এশিয়াটিক সোসাইটি।

Leech, Geoffrey. (1992). "Corpora and Theories of Linguistic Performance." In Svartvik (1992). 105-22.

MacWhinney, Brian. (2000). *The CHILDES Project : Tools for Analyzing Talk*. 3rd edn., vol. 1: *Transcription*

Format and Programs, vol 2 : *The Database*. Mahwah, NJ: Erlbaum.

Marcus, Mitchell P., Beatrice. Santorini, and Mary Ann. Marcinkiewicz. (1993). "Building a Large Annotated Corpus of English: The Penn Treebank", *Computational Linguistics* 19. 314-30.

Mayer, Charles, F. (2004). *English Corpus Linguistics: An Introduction*. Cambridge : Cambridge University Press.

McEnery, T. and Wilson, A. (1996). *Corpus Linguistics*. Edinburgh : Edinburgh University Press.