



हिंदी-मराठी अनुवाद में शाब्दिक अस्पष्टता

प्रकाश अभिमन्यु कांबळे

panka2007@gmail.com

शोधार्थी, जेएनयू, दिल्ली

आधुनिकीकरण के युग में भाषाविद यह नहीं भूल सकते की भारत एक बहुभाषिक देश है। बहुभाषिकता की समस्या को अनुवाद के माध्यम से दूर किया गया, लेकिन आई.टी के युग में भाषा का मुकाबला मशीन से है। मशीनी अनुवाद के क्षेत्र में कई समस्याएं हैं। जिसमें से एक प्रमुख समस्या समान उच्चारण वाले भिन्नार्थक शब्द (Homonymy) हैं। इस समस्या पर वैश्विक स्तर पर विचार किया जा रहा है। आवश्यकता है कि भारतीय भाषाओं में भी इस विषय पर चर्चा हो। Homonymy की समस्या केवल मशीनी अनुवाद में ही नहीं मानव द्वारा किए जा रहे अनुवादों में (Human Translation), आशु अनुवादों (Interpreting) में, नई जुबान सीखते वक्त (Language Acquisition), सामान्य बुद्धि के छोटे बच्चों एवं कृत्रिम बुद्धि विशेषज्ञों (Artificial Intelligence specialists) को भी Homonymy शब्दों के अस्पष्टता की समस्या ने परेशानी में डाला है। सबसे पहले यह जान लें कि समान उच्चारण वाले भिन्नार्थक शब्द तत्सम, तद्भव या देशी शब्दों के ही शब्द-युग्म हैं। यह शब्द हिंदी में प्रयुक्त विदेशी शब्दों और हिंदी शब्दों के भी रूप होते हैं। समध्वनिय (homophones) भिन्नार्थक शब्द-युग्म का उच्चारण समान रूप से होता है, लेकिन समध्वनिय शब्दों की वर्तनी असमान भी हो सकती है। समान उच्चारण वाले

भिन्नार्थक शब्द में कई बार वर्तनी भी समान होती है। देवनागरी लिपि ध्वनिमूलक होने के कारण वर्तनी में आने वाली भिन्नता से उच्चारण में भी भिन्नता आ जाती है।

Homonymy शब्दों की समस्या के विसंदिग्धिकरण (Disambiguation) सूत्र की ओर बढ़ने से पहले ऐसे शब्दों की व्याख्या और इन शब्दों का विस्तृत परिचय कराना जरूरी है। जिससे समस्या की मूल गहराई को समझा जाए। व्याख्या -“In linguistics, a Homonym is one of a group of words that share the same pronunciation but have different meanings, and are usually spelled differently or same.” [1]

उपरी व्याख्या से हमें पता चलता है कि “समान उच्चारण वाले भिन्नार्थक शब्द-युग्म है, जिनका उच्चारण प्रायः समान होता है, किंतु उनके अर्थ भिन्न होते हैं और कई जगह वर्तनी समान या अलग भी हो सकती है।” हिंदी में कुछ ऐसे शब्द हैं, जिनका उच्चारण समान है। वर्तनी भिन्न है और और अर्थ समान है। लेकिन इन शब्दों की संख्या काफी कम है। जिसे हम अपवाद भी समझ सकते हैं। सबसे पहले Homonymy शब्दों को कितने भागों में विभाजित किया जा सकता है इसपर विचार किया जाए। जिससे इन शब्दों का ज्यादा से ज्यादा संशोधित रूप हमारे सामने आ जाए, जिससे आगे बढ़ने में आसानी होगी। Homonymy शब्दों को निम्न भागों में विभाजित किया जा सकता है।

1. Homographs
2. Homophones
3. Heteronyms
4. Polysemes
5. Capitonyms

	Same Pronunciation	Different Pronunciation
Same Spelling	  	 
Different Spelling		

 <u>Homonym</u>	 <u>Homophone</u>
 <u>Homograph</u>	 <u>Heteronym/ Heterophony</u>

उच्चारण (Pronunciation), समध्वनि और भिन्नार्थकता के निकषों पर शब्द-युग्मों को हम विभाजित कर सकते हैं। जिससे समान उच्चारण वाले भिन्नार्थक शब्दों की असमानता (inequality) को आसानी से समझ सकते हैं।

क्र.	वर्तनी	उच्चारण	अर्थ
1	समान वर्तनी	समान उच्चारण	उच्चारण
2	भिन्न वर्तनी	समान उच्चारण	भिन्न अर्थ
3	भिन्न वर्तनी	भिन्न उच्चारण	समान अर्थ

शब्दकोश के माध्यम से किये जाने वाले शब्दों के आशय को व्याकरणिक कोटियों या शब्दों की व्युत्पत्ति के आधार पर निर्धारित किया जाता है। लेकिन समस्या तब उत्पन्न होती है, जब कम्प्यूटर केवल शब्दकोश में दिए गये शब्दों को शब्द संचयन (Lexical Data) में दिये गये निर्गमण (Input) क्रम के अनुसार ही शब्दों का वाचन (read) करता है, जिसका निवेशन (output) आशय (senses) वाक्य में प्रयुक्त आशय के समान नहीं होता। शब्दकोश में शब्द समूह आशय की इस गलती के कारण स्रोत भाषा पाठ से आये शब्दों को लक्ष्य पाठ में सही अर्थों में अंतरण

(Transfer) करने में सफल नहीं हो पाता है। शब्दकोश समूह आशय से मिले निवेशन (output) के प्रतिशत को अधिक बढ़ाने के लिए निवेशन कलन विधि (Algorithmic Output) की आवश्यकता है। जिसे Homonymy शब्द विभाजन के अनुसार विभाजित शब्दों और व्याकरण से पूरा किया जा सकता है। लेकिन अनुवादकों एवं आम लोगों से यह सवाल बार-बार उठाया जाता रहा है कि “मौजूदा सर्वोत्तम प्रोग्राम के माध्यम से भी विकल्पों की एक शृंखला ही प्रस्तुत की जा सकती है जिससे सर्वश्रेष्ठ अथवा बेहतर का चयन अनुवादक को ही करना होगा।” [2] कलन विधि पर उठे इस सवाल का जवाब भी समान उच्चारण वाले भिन्नार्थक शब्दों की समस्या को अंजाम देकर दूर किया जा सकता है।

कंप्यूटर की कई सीमाओं और समस्याओं पर विचार कर ज्यादा से ज्यादा उपयोग में लाने से पहले किसी एक जुबान पर पूरी सफलता के साथ काम किया जाए। उसके बाद ही दूसरी जुबानों पर उसका इस्तेमाल किया जाए। इसी के साथ इसपर भी ध्यान दिया जाए की बड़ी समस्याओं को छोटे-छोटे भागों में विभाजित कर उस पर संशोधन किया जाए। छोटी समस्याओं के समाधान से बड़ी समस्या का समाधान हो जाएगा। यह पेपर इसी प्रकार की मशीनी अनुवाद में उपन्न हुई अस्पष्टता (Ambiguity) की बड़ी समस्या में से छोटी समस्या का समाधान खोजने का प्रयास है। जिसे निम्न रूप से हिंदी से मराठी अनुवाद को ध्यान में रखा है। हिंदी स्रोत भाषा होने के कारण अधिक ध्यान हिंदी पर दिया जाना चाहिए, क्योंकि Homonymy की समस्या पहले स्रोत भाषा में उत्पन्न होती है, जिसका निराकरण भी स्रोत भाषा में ही होना चाहिए। लक्ष्य भाषा में केवल निर्माण होने वाली समस्या के निराकरण के अर्थ का अनुवाद ही करना होगा।

क्र.	श.क्र.	हिंदी शब्द	व्या.को.	हिंदी अर्थ	मराठी अर्थ
------	--------	------------	----------	------------	------------

1	1	कुल	नाम<पुल्लिंग<एक वचन	कुटुंब, कुनबा	कुळ (नाम<पुलिंग<एक व.
	2	कुल	नाम<पुल्लिंग<एक वचन	जोड़	एकूण< नाम<पुल्लिंग<एक वचन
2	1	अचल	विशेषण	पर्वत	पर्वत <विशेषण
	2	अचल	विशेषण	स्थिर	स्थिर <विशेषण
3	1	कोटि	नाम< स्त्री.	दर्जा	दर्जा <नाम <स्त्री
	2	कोटि	नाम< स्त्री.	करोड़	करोड़ <नाम <स्त्री

समान उच्चारण- समध्वनीय – समान वर्तनी वाले- भिन्नार्थक शब्द

शब्द अकृतियों के माध्यम से ही बनते हैं और ये आकृतियाँ ही समान उच्चारण वाले भिन्नार्थक शब्दों की अस्पष्टता के निराकरण में सहायभूत होंगी। प्रत्येक शब्द की एक भिन्न आकृति होती है और आकृति के साथ ही अर्थ भिन्न होते हैं। प्रथम कार्य यह होना चाहिए की आकृति के माध्यम से अर्थ तक पहुँचे, क्योंकि अर्थ ही भाषा की आत्मा है। यह कार्य असमान आकृतियों के लिए सहायभूत होगा लेकिन समान आकृतियों वाले शब्दों पर यह नियम लागू नहीं होगा। समान आकृतियों वाले शब्दों के लिए कलन विधि के द्वारा ही इन शब्दों का निराकरण हो पायेगा। समान उच्चारण और समान ध्वनि वाले शब्दों की संख्या कम होने के कारण यह कार्य अधिक कष्टकर नहीं होगा। जिससे कलन विधि से किया गया कार्य अधिक संशोधन पूर्ण होगा।

क्र.	श.क्र.	हिंदी शब्द	व्या.को.	हिंदी अर्थ	मराठी अर्थ
1	1	दिन	Noun Mas.	दिवस	दिवस (नाम.पु)

	2	दीन	Adj.	गरीब	गरीब (वि.)
2	1	बलि	Noun Mas.	बलिदान	बलि (नाम.पु)
	2	बली	Adj.	बलवान	शक्तिशाली (वि)
3	1	बाजि	Noun Mas.	घोड़ा	घोडा (नाम.पु)
	2	बाज़ी	Noun Fem.	दाँव	बाजी (नाम.स्त्री)

समान उच्चारण – समध्वनीय – भिन्न वर्तनी – भिन्न अर्थ

हिंदी में वर्तनी की भिन्नता से उच्चारण में भिन्नता आ जाती है। अविराम-अभिराम; सम-शम; सूर-शूर, आसन-आसन्न आदि शब्द-युग्म उच्चारण एक जैसे हैं। अंग्रेजी में अवश्य ही शब्दों के ऐसे जोड़े मिलते हैं, जिनकी वर्तनी भिन्न होती है, परन्तु उच्चारण समान होता है। जैसे – Wood-would; Know-No; Bell-Bail; pain-pen; Sun-Son इत्यादि। अंग्रेजी में ऐसे शब्दों को “होमोफोन्स” कहा गया है। [3] यद्यपि हिन्दी में समध्वनि भिन्नार्थक शब्द सही अर्थ में होमोफोन्स नहीं है, फिर भी इन शब्दों का किन्हीं अन्य शब्दों से समानता हो सकती है, तो वह होमोफोन्स से ही। होमोफोन्स भले ही ध्वन्यात्मकता से पहचाने जाते हों लेकिन हिंदी में वर्तनी भिन्नता के कारण यह समस्या अधिक महत्त्वपूर्ण नहीं रह जाती। इसका निराकरण काफी आसानी से किया जा सकता है। और वह आसान तरीका है वर्तनी के आकार के अनुसार कलन विधि (Algorithm) के द्वारा। (ऐसे नाम जो Homonymy शब्दों की तरह अस्पष्ट होते हैं, जिनका उच्चारण समान होता है और कई बार भिन्नार्थक होते ऐसे शब्दों की ध्वनि और वर्तनी समान होती है।)

क्र.	श.क्र.	हिंदी शब्द	व्या.को.	हिंदी अर्थ	मराठी अर्थ
1	1	देव	Noun Mas.	भगवान	देवता (नाम.पु.)
	2	देव	Noun	नाम	नाव (नाम)
2	1	धवल	Adj.	सफेद	पांढरा(वि.)
	2	धवल	noun	नाम	नाव (नाम)
3	1	दिवाकर	Noun Mas.	सूर्य	सूर्य (नाम.पु)
	2	दिवाकर	noun	नाम	नाव (नाम)

समध्वनि भिन्नार्थक नामों के कारण अस्पष्टता की समस्या सभी भाषाओं में समान रूप से मिलती है। विशेष रूप से ऐसे नाम, जो आम बोलचाल में प्रयुक्त किए जाते हैं या भाषा में काफी आसानी से उपयोग में आते हैं। लेकिन छोटे बच्चों और मशीन को ऐसे शब्दों से काफी परेशानी का सामना करना पड़ता है। विद्वानों ने ऐसे नामों को Homonymy के अंतर्गत नहीं रखा है। इस कारण इस पर विचार नहीं किया जा रहा है।

क्र.	श.क्र.	हिंदी शब्द	व्या.को.	हिंदी अर्थ	मराठी अर्थ
1	1	टेस्ट	Noun	स्वाद	चव (नाम)
	2	टेस्ट	Verb transitive	इम्तहान	परिक्षा (नाम)
2	1	इंटरेस्ट	Noun	दिलचस्पी	रुचि (नाम)

	2	इंटेरेस्ट	Verb transitive	ब्याज	व्याज (क्रि.सकर्मक.)
3	1	कंडिशन	Verb transitive	हालत	परस्थिति (क्रि.सकर्मक.)
	2	कंडिशन	Verb transitive	प्रतिबंध	नियम (क्रि.सकर्मक.)

समान उच्चारण – समध्वनि/विषम-ध्वनि – समान वर्तनी-भिन्न अर्थ

भाषा के विकास के लिए दूसरी भाषाओं से शब्द लेने की परंपरा का एक नतीजा यह हो रहा है कि Homonymy शब्दों की संख्या में बढ़ोतरी हो रही है, जो भाषा में अनेकार्थकता निर्माण करती है। हिंदी में कई भाषाओं के शब्द हैं। उच्चारण और ध्वनि समान होने के बावजूद अर्थ में भिन्नता पाई जाती है। ऐसे शब्दों को डॉ. त्रिभुवन ओझा ने तीन भागों में विभाजित किया है। जिससे अंशतः अर्थ की स्पष्टता हो सकती है।

1. इस कोटि में वे शब्द आते हैं जो अपने मूल उच्चारण के साथ हिन्दी शब्दों के समरूप हो गए हैं। जैसे :- जून (June), फूट (Foot), बूट (Boot)
2. इस कोटि में वे शब्द आते हैं जो साधारण ध्वनि-परिवर्तन के बाद हिन्दी शब्दों के समरूप हो गये हैं। जैसे:- चौक (Chalk), आकदमी (academy)
3. इस कोटि में वे शब्द आते हैं जो अंग्रेजी शब्दों के तद्भव रूप कहे जा सकते हैं।
जैसे :- बम (bomb)' काग (cork) ऐसे शब्दों को अर्थ विशेषता के साथ वर्गीकृत कर व्याकरणिक नियमों से स्पष्ट किया जा सकता है।

निम्नलिखित शब्दों को किस कोटि में रखा जाए यह एक समस्या है, क्योंकि यह शब्द मानक नहीं होते। यह शब्द रोज-मर्रा के जीवन में नियमित रूप से आते हैं, जिसका उपयोग अशिक्षित या गवाँरू

लोग अधिक करते हैं। इन शब्दों की वर्तनी भिन्न होती है, उच्चारण होमोफोन्स की तरह होता है लेकिन यह शब्द भिन्नार्थक नहीं होते। यह शब्द प्रामाणिक नहीं है। होमोनिम शब्द शब्दकोशों में भी बहुत कम मिलते हैं। इस लिए इन शब्दों पर विचार नहीं किया जा रहा। जैसे :-

क्र.	श.क्र.	हिंदी शब्द	व्या.को.	हिंदी अर्थ	मराठी अर्थ
1	1	प्रतिकार	Noun Mas.	प्रतिशोध	प्रतीशोध (ना.पु.)
	2	प्रतीकार	Verb transitive	प्रतिशोध	प्रतीशोध (क्रि.अ.)
2	1	टेबल	Noun	तालिका	तक्ता (नाम. नपुसक लिंग)
	2	टेबुल	Noun	तालिका	तक्ता (नाम. नपुसक लिंग)
3	1	टेबल	Noun	मेज़	टेबल (नाम. नपुसक लिंग)
	2	टेबुल	Noun	मेज़	टेबल (नाम. नपुसक लिंग)
4	1	दुलहा	Noun Mas.	वर	नवरा (नाम)
	2	दुल्हा	Noun Mas	वर	नवरा (नाम)

भाषा-प्रौद्योगिकी में समान उच्चारणवाले सभी शब्दों को अन्तर्विष्ट (contains) कर एक साथ विसंदिग्धिकरण (Disambiguation) की सभी पद्धतियों और रूपरेखाओं (scheme) के पारस्परिक क्रिया से भी दूर किया जा सकता है।

भाषा सृजनशीलता में नामकरण विधि एक सृजनता है। जिसमें सभी उद्देश्य असमान रूप से व्यक्तिवाचक नाम के समान अभिज्ञेय (identifiable) होते हैं, लेकिन “सेव” और “कुर्सी” जैसे वस्तु वाचक नामों के समान नहीं होते। कुछ उद्देश्यों को समान विषयों के साथ वर्गीकृत किया जाता है। हर उद्देश्य एक विशेष नाम में वर्गीकृत होता है। ऐसे हालात में स्रोता और वक्ता दोनों के बीच का संप्रेषण शून्य स्थिति में होता जाएगा। जिससे समान उच्चारण वाले शब्दों में अस्पष्टता (अराजकता) नहीं रह पायेगी। जिससे स्रोता और वक्ता के शब्दों के अर्थों का उद्देश्य स्वयं अपने साथ अपरिचित या एकमात्र अर्थों को स्पष्ट करेगा। सबसे महत्वपूर्ण कार्य यह होना चाहिए homonymy शब्दों की अस्पष्टता को दूर करने के लिए विशेष रूप से बनाए गए भाषिक संसाधनों का अधिक से अधिक उपयोग हो। जैसे:- कार्पस, टॅगर, रूप-विश्लेषक और विशेष शब्दकोश जिसमें केवल Homonymy शब्दों का संचयन हो। जिससे कार्य में भी गति आ सके। कार्पस में केवल Homonymy शब्दों की जानकारी शब्द संचयन ही न होकर शब्दों की उपयोगिता पर भी लक्ष्य केंद्रित किया जाए। जैसे :-

1. आना (Noun)=(हिंदी) <भिखारी> <का> <कटोरा> <आठ> <आने> <और> <चार> <आने> <से> <भरा> <हुआ> <था><।> (मराठी) <भिका-याची> <वाटी> <चार> <आने> <आणि> <आठ> <अन्यांनी> <भरली> <होती> <.>
2. आना (Verb) (हिंदी) – <उसका> <आना> <मेरे> <लिए> <कितना> <सुखद> <था> <।>
(मराठी) – <त्याच> <येणं> <माझ्या> <साठी> <किती> <सुखद> <होतं> <.>

कार्पस में दिये गये वाक्यों को फिर से Homonymy शब्दकोश की सहायता से संशोधित किया जा सकता है, जिससे अर्थ की स्पष्टता होगी। रूप-विश्लेषक का उपयोग रूप को विश्लेषित करने में सहायभूत होगा। जब समान वर्तनी और भिन्न वर्तनी के शब्दों को आसानी से विश्लेषित कर सकता है। जिसपर समध्वनि और उच्चारण पर अधिक टिका हुआ है।

इस पेपर में यह ध्यान रखा गया है कि सिर्फ अर्थ-स्थानांतरण का सरलीकरण न होने पाये। यह प्रयास है कि एक ऐसे यंत्र-रचना (mechanism) का निर्माण हो जो स्वयं ही न्यूनतम अर्थ को संप्रेषित करने में समर्थ हो। प्रस्तुत पत्र हिंदी-मराठी मशीनी अनुवाद के संदर्भ में शाब्दिक अस्पष्टता का चिह्न और निराकरण खोजने का प्रयास है।

संदर्भिका

[1] <http://en.wikipedia.org/wiki/Homonym>

[2] कंप्यूटर अनुवाद:प्रयोग और विधि – प्रो.रीतारानी पालीवाल, अनुवाद पत्रिका (कंप्यूटर अनुवाद विशेषांक -२) अप्रैल-जून २००४ पेज – ५६

[3] हिंदी में अनेकार्थता का अनुशीलन – डॉ.त्रिभुवन ओझा करीम सिटी कॉलेज, जमशेदपुर, विश्वविद्यालय प्रकाशन, वाराणसी १६६४ पेज – ७७

- A cross-situational learning algorithm for damping homonymy in the guessing game, – Joachim De Beule, Bart De Vylder and Tony Belpaeme, Vrije Universiteit Brussel, Belgium, University of Plymouth, United Kingdom



- Integral Technology of Homonymy Disambiguation in the text mining system “LOTA”, – Olga Nevzorova, Vladimir Nevzorov, Julia Zin’kina, Nicolay Pjatkin
- Particle Homonymy and Machine Translation, – K6roly F&bricz, JATE University of Szeged, Egyetem u. 2. Hungary – 6722
- Children’s difficulty in learning homonyms*, – MARTIN J. DOHERTY, Department of Psychology, University of Stirling
- Native and L2 processing of homonyms in sentential context, – Kerrie E. Elston-Gu” tler*, Angela D. Friederici, Max Planck Institute of Human Cognitive and Brain Sciences, Leipzig, Germany
- Particle Homonymy and Machine Translation, Kroly F&bricz, JATE University of Szeged, Egyetem u. Hungary – 6722
- Learning Form-Meaning Mappings in Presence of Homonymy: a linguistically motivated model of learning inflection, Katya Pertsova, University of California Los Angeles
- प्रामाणिक हिंदी शब्द-रचना एवं वर्तनी प्रकाश –
- हिंदी में अनेकार्थता का अनुशीलन – डॉ.त्रिभुवन ओझा करीम सिटी कॉलेज, जमशेदपुर, विश्वविद्यालय प्रकाशन, वाराणसी
- कंप्यूटर अनुवाद: प्रयोग और विधि-प्रो.रीतारानी पालिवाल, अनुवाद पत्रिका (कंप्यूटर अनुवाद विशेषांक-२) अप्रैल-जून २००४ पेज-५६
- <http://en.wikipedia.org/wiki/Homonym>
- <http://assortedmaterial.googlepages.com/EnglishIndex.html>



- <http://www.tribuneindia.com/2000/20000819/windows/roots.htm>

Citation: कांबळे, प्रकाश अभिमन्यु (2010). हिंदी-मराठी अनुवाद में शाब्दिक अस्पष्टता, HindiTech: A Blind Double Peer Reviewed Bilingual Web-Research Journal, 1 (4), 49-61. URL: <https://hinditech.in/hindi-marathi-anuvad-men-shabdik-aspashtata/>