अध्याय : 2

आर्थी विश्लेषण एवं इसका तकनीकी पक्ष

(Semantic Analysis and its Technical Aspect)

2.0 प्रस्तावना

2.1 अर्थ

2.2 अर्थबोध और बिंदनिर्माण

2.3 मानव मस्तिष्क में बिंदबूं की व्यवस्था

2.4 ज्ञान निरूपण : अर्थ-व्यवस्थापन की तकनीकी विधि

2.5 ज्ञान निरूपण /अर्थ व्यवस्थापन की कुछ विधियाँ

2.6 ज्ञान निरूपण और आर्थी विश्लेषण

2.7 आर्थी विश्लेषण: अन्य पक्ष

2.8 उपसंहार
आर्थी विश्लेषण एवं इसका तकनीकी पक्ष
(Semantic Analysis and its Technical Aspect)

2.0 प्रस्तावना

भाषा व्यवस्था के दो पक्ष हैं- अभिव्यक्ति पक्ष और कथ्य पक्ष। इनमें अभिव्यक्ति पक्ष का संबंध जहाँ ‘ध्वनि’ से है वही कथ्य पक्ष का संबंध ‘अर्थ’ से है। भाषा के संदर्भ में कहा जाता है कि यह संप्रेषण का एक माध्यम है। इसे ही दूसरे शब्दों में कहा जा सकता है कि ध्वनियों के माध्यम से अर्थ को संप्रेरित करने वाले एक व्यवस्था है। ध्वनि एवं अर्थ के संबंधों को देखने पर एक वात स्पष्ट हो जाती है कि जहाँ ‘ध्वनि’ है वहाँ अर्थ नहीं है और जहाँ अर्थ है वहाँ ‘ध्वनि’ नहीं है।

इन दो छोरों को भाषा अपनी विभिन्न संरचनात्मक इकाइयों के माध्यम से जोड़ने कार्य करती है।

किसी संप्रेषण या भाषा के अर्थ पक्ष से नज़रों इकाइयों एवं घटनाओं का विश्लेषण ‘आर्थी विश्लेषण’ है। इसमें भाषिक इकाइयों एवं उनके घटकों को उनमें निहित अर्थ और भाषिक संरचनाओं के निर्माण में लगे अर्थपरक प्रतिवर्तों का अध्ययन विश्लेषण किया जाता है। आर्थी विश्लेषण को समझने तथा इसकी व्याख्या करने के दो पक्ष हैं- पारंपरिक भाषावैज्ञानिक पक्ष तथा आधुनिक तकनीकी पक्ष। पारंपरिक भाषावैज्ञानिक पक्ष में विभिन्न भाषिक इकाइयों- स्वरिक, रूपम, शब्द, पदवंसं, वाक्य एवं प्रक्ति तथा इनमें अर्थ के स्वरूप एवं विस्तार का विवेचन किया जाता है। जैसा कि हम जानते हैं, भाषिक इकाइयों में अर्थ का विस्तार रूपम से प्रक्ति से तक पाया जाता है। इस वात और है कि प्रत्येक स्तर पर इसका स्वरूप अलग-अलग होता है। इसके विपरीत यदि आधुनिक तकनीकी पक्ष की बात करें तो यह मशीनों में प्राकृतिक भाषाओं के प्रयोग के सापेक्ष विकसित हुआ है जिसके मुख्य क्षेत्र प्राकृतिक भाषा संसाधन (Natural Language Processing) एवं कृत्रिम बुद्धि (Artificial intelligence) हैं। इनमें प्राकृतिक भाषाओं में प्रास संरचनाओं एवं उनसे जुड़े अर्थों को इस प्रकार संरचित करने का प्रयास किया जाता है कि एक मशीन भी अभिव्यक्ति के माध्यम से अर्थ को समझ सके। इसके लिए सभी संरचनात्मक इकाइयों एवं उनसे जुड़े कथ्य को पराक्रिक अभिव्यक्तियों के रूप में ढालने की आवश्यकता होती है।
2.1 अर्थ

अपने प्रकार की हत्ति से अर्थ संप्रेषण का वास्तविक एवं बादश पक्ष तथा भाषा का मानसिक संकल्पनात्मक पक्ष है। मानव मस्तिष्क में इसकी उपस्थिति संकल्पनात्मक इकाईयों के संग्रह के रूप में मानी गई है। इसे व्याख्यात करने के लिए 'आग्नि एवं रिचर्ड्स' (1993) द्वारा दी गई अर्थ की परिभाषा का उल्लेख प्रारंभिक है जिसे अनूठित रूप में दिया जा रहा है:

“अर्थ यह मानसिक तत्व है जो एक ओर घटनाओं और विषयों तथा दूसरी ओर उनके लिए प्रयोग में लाए जाने वाले प्रतीकों और शब्दों के बीच होता है।”

अर्थ को संकेतार्थ कहते हुए इसे उपर्युक्त विद्वानों द्वारा दिए गए सूचन के त्रिकोण को देखा जा सकता है:

![संकेतित वस्तु]

अर्थ के मस्तिष्क में स्थित होने की दो स्थितियाँ हैं: शब्दार्थ स्तर पर संकल्पना के रूप में और वाक्यात्मक स्तर पर एक से अधिक संकल्पनात्मक इकाईयों के परस्पर संयोजित होने की स्थितियों के रूप में। जब किसी शब्द का उच्चारण किया जाता है तो उसकी संकल्पना हमारे मस्तिष्क में उभरती है, जैसे- यदि ‘लड़का’, ‘कौआ’ और ‘उड़ना’ शब्दों का उच्चारण किया जाए तो इनकी संकल्पना हमारे मस्तिष्क में उभरती है। अब यदि इनका संयोजन करते हुए दो वाक्यों का निर्माण किया जाए:

1. कौआ उड़ता है।
2. लड़का उड़ता है।

तो इनमें प्रथम वाक्य तो ठीक है किंतु द्वितीय वाक्य में आए घटक शब्दों की संकल्पनाएँ परस्पर संयोजित नहीं हो पा रही हैं। इसका कारण बाद जगत में ‘लड़का’ के अंदर ‘उड़ने’ की क्षमता का अभाव होता है।
इस प्रकार की बातों को समझने तथा व्याख्यातित करने के लिए अर्थ को समझने एवं इसे उत्तिच्च प्रकार से निरूपित करते हुए विवेचित करने की आवश्यकता पड़ती है। यह आवश्यकता तब और महत्वपूर्ण हो जाती है जब इस प्रकार के ज्ञान को मशीन में स्थापित करता हो। क्योंकि मशीन में तो स्वतः से कुछ भी करने की थोड़ी भी क्षमता नहीं है। यह कार्य मनुष्य द्वारा कैसे किया जाता है? इस पर थोड़ी चर्चा कर लेना अप्रासंगिक नहीं होगा।

2.2 अर्थबोध और विविधनिर्माण

भाषिक इकाइयों (मूलतः ‘वाक्य’) का व्यवहार होने पर अर्थ को ग्रहण करना ही अर्थबोध है।

उदाहरण के लिए एक वाक्य को देखें:

“उस बोलता का पानी दीजिए।”

जब वका द्वारा इस वाक्य का उच्चारण किया जाता है तब श्रोता उठकर पास में रखे किसी बोलते के पानी को उठकर वका को प्रदान करता है। ‘किसी बोलता का पानी उठकर देना’ एक भौतिक क्रिया है जिसे उपयुक्त वाक्य को सुनने के पश्चात् श्रोता संपन्न करता है। अतः कहा जा सकता है कि यह भौतिक क्रिया इस वाक्य का अर्थ है। अतः एक वात देखी जा सकती है कि वका द्वारा इस वाक्य के उच्चारण एवं श्रोता द्वारा ‘पानी दिए जाने’ या ‘पानी देने के लिए तत्पर/तैयार होने’ के से पूर्व श्रोता के मस्तिष्क में एक क्रिया संपन्न हुई। इस क्रिया को ही अर्थबोध कहते हैं।

अब एक विचारणीय प्रश्न यह है कि वका द्वारा भाषिक इकाइयों, जैसे- शब्द, वाक्य आदि का उच्चारण या व्यवहार करने पर श्रोता को उसका अर्थबोध कैसे होता है। निश्चय ही यह एक मानसिक प्रक्रिया है। इस स्तर पर जो मानसिक प्रक्रिया पृथ्वभूमि में काम करती है उसे ‘विविधनिर्माण’ कहते हैं।

जैसा कि उस पर ‘अर्थ’ उपशिरक वक्ते अंतर्गत चर्चा करते हुए इसे एक मानसिक संकेतार्थ के रूप में बदला गया है। यही तत्व उच्चारित शब्द या वाक्य को बाह्य जगत की संकेतित वस्तु से जोड़ता है। इस तत्व के मस्तिष्क में स्थापित होने की प्रक्रिया इस प्रकार है:

शब्द के साथ ऐच्छिक (समाज द्वारा स्वीकृत) रूप से संबंधित किए गए अर्थ की पूरी व्याख्याता का ज्ञान मनुष्य के अनुभव से होता है। जन्म के पश्चात् जैसे-जैसे मानव शिशु में चेतना का विकास होता है वैसे-वैसे वह परिवेश की वस्तुओं से परिचित होता जाता है। इससे अर्थ रूपी विकास का निर्माण उसके मस्तिष्क में होने लगता है। अर्थात् की इस प्रक्रिया में मानव शिशु अनावास और अनजाने रूप में अर्थ के भेदक तथा अभेदक तत्वों का विश्लेषण करता है। इस प्रकार प्रत्येक वस्तु, कार्य और संकल्पना के लिए मस्तिष्क में अलग-अलग विचार अंकित हो जाते हैं। इसी प्रक्रिया को पारंपरिक रूप से विविधनिर्माण कहा गया है।
बिंदुर्निर्माण की प्रक्रिया का दूसरा पक्ष 'भाषाई रूप' (linguistic forms) का निर्माण है। भाषा हमारे लिए लौटक जगत की वस्तुओं और कार्यों को समझने, विचार करने तथा इन्हें अभिव्यक्त करने का माध्यम है। अत: प्रत्येक बिंदु के लिए एक भाषिक रूप (मुख्यतः शब्द) का होना आवश्यक है। लौटक जगत की वस्तुओं, कार्यों और मूर्त/अमूर्त संकल्पनाओं को देखने पर या उनका बोध होने पर उन्हें समझने और अभिव्यक्त करने के लिए शब्दों की आवश्यकता होती है। ये शब्द (स्वयं समूह और उनकी व्यवस्था) हमें अपने समाज और परिवेश से प्राप्त होते हैं। धीरे-धीरे ये शब्द अभिव्यक्ति के रूप में उस बिंदु-विशेष या तत्त्व रूप से जुड़ जाते हैं कि शब्द (भाषिक रूप) के आने पर अर्थ (बिंदु) और अर्थ के आने पर शब्द मस्तिष्क में स्वयं ही आ जाता है। शब्द द्वारा वस्तुबोध होने की प्रक्रिया को ‘अर्थवोध' या ‘संकेतग्रह' कहा गया है।

2.3 मानव मस्तिष्क में बिंदुओं की व्यवस्था

इस प्रकार हम देखते हैं कि भाषा में एक तरफ बाह्य वस्तु, कार्य या भाव के आधार पर अर्थ (बिंदु) का समावेश होता है और दूसरी तरफ उसके लिए मस्तिष्क में एक भाषिक रूप (शब्द) स्थापित होता है जो हमें समाज से प्राप्त होता है। मस्तिष्क में निर्मित होने वाला बिंदु अनुभूति पर आधारित होता है। यह अनुभूति इंद्रियजन्य होती है, अत: बिंदु के अंतर्गत किसी वस्तु का चित्र, उसके गुण या विशेषताएं, लक्षण, कल्पना, आभास और विचार आदि सभी आ जाते हैं। यहाँ पर एक बात महत्वपूर्ण है कि अर्थविश्लेषण में संकल्पनात्मक संरचनाओं (conceptual structures) की वस्तु का गई है क्योंकि अमूर्त संज्ञाओं की कोई इंद्रियजन्य अनुभूति नहीं होती। यहाँ पर ‘शब्द + संकल्पना' (लक्षण समुच्चय) के रूप में इन्हें भी बिंदु के अंतर्गत समाहित कर लिया जा रहा है। दूसरी तरफ भाषिक रूप अपनी व्यवहारिक इकाइयों के प्रतिनिधि के रूप में अमूर्त रूप से मानव मस्तिष्क में स्थापित होते हैं।

अब एक प्रश्न यह उठता है कि ये तत्त्व मानव मस्तिष्क में किस प्रकार संगठित होते हैं? विभिन्न विद्वानों द्वारा इस संबंध में अपने-अपने दृष्टिकोण से विचार किया गया है। मानव मस्तिष्क में भाषा किस प्रकार संगठित तथा व्यवस्थित है इसके केवल अनुमान ही लगाए जा सकते हैं क्योंकि ऐसी कोई व्यवस्था विकसित नहीं हो पाई है जो मस्तिष्क में भाषा एवं आर्थिक बिंदुओं को चित्रित कर सके। वैसे बिंदुओं के व्यवस्थित रूप से संचित होने की वस्तु की जाए तो एक बात तो स्पष्ट है कि किसी बाह्य वस्तु के निरूपित होने की मुख्यतः दो विधियों हैं: प्रथम- चित्र का अंकित होना एवं, द्वितीय- विशेषताओं, कार्यों आदि का अंकित होना, जैसे: किसी व्यक्ति के मस्तिष्क में गेंद का बिंदु कुछ प्रकार से होगा:

(क) चित्र के रूप में (गेंद का चित्र)
(ख) एक ऐसी वस्तु के रूप में जो (क) गोल होती है, (ख) आकार में छोटी होती है (लगभग मुठ्ठी भर की), (ग) जिससे खिलते हैं आदि।
दोनों के माध्यम से ही ‘गेंद’ को समझा जा सकता है। इसी प्रकार से अन्य वस्तुओं एवं संकल्पनाओं के भी विंब मानव मस्तिष्क में स्थित होते हैं।

वस्तुओं के विंबों का मानव मस्तिष्क में स्थापित होना एक बात है। उन विंबों का परस्पर एक दूसरे से संबंधित होना होना दूसरी बात। इन विंबों के विंब रूपों एवं उनके संबंधों को ‘अर्थ के विंब रूप’ कहते हुए अर्थविज्ञान के अंतर्गत चर्चा देखी जा सकती है। इनमें से कुछ प्रमुख को यहाँ पर संकेत में लिया जा रहा है:

(क) अवनामिता और अधिनामिता (Hyponymy and Hypernymy) : जब किसी शब्द के अर्थ का उससे बड़े वर्ग में समावेश किया जाता है तो इस इकाई एवं इससे बड़ी इकाई के बीच प्रास होने वाले संबंध में बड़ी इकाई से छोटी इकाई के क्रम को अवनामिता एवं विपरीत क्रम को अधिनामिता कहते हैं। इसे ‘जाति-सदस्य संबंध’ या ‘वर्ग-सदस्य संबंध’ भी कहा जा सकता है। इसमें छोटी इकाई या वर्ग को अवनामी (hyponym) और बड़ी इकाई या वर्ग को अधिनामी (hypernym) कहते हैं। इसे निम्नलिखित प्रकार से देखा जा सकता है:

(ख) विलोमता (Antonymy) : यदि दो शब्दों के अर्थ में परस्पर विरोध या विपरीततमत्कर्ता हो तो ऐसे शब्दों को विलोम शब्द और इस स्थिति को ‘विलोमता’ कहते हैं। वर्तमान में इस संबंध को नए रूप दिखाया जा सकता है।

अ) विरोधात्मकता (contradictoriness): इसमें एक के होने का अर्थ है दूसरे का नहीं होना, जैसे: जीवन-मृत्यु।

आ) विपरीततमता (oppositeness) : इनमें एकदम विरोध प्रास नहीं होता, जैसे: धनी-गरीब।

इ) पूरकता (complementariness): इसमें दोनों मिलकर एक इकाई या चक्र का निर्माण करते हैं, जैसे: दिन-रात।

ई) विपर्यय (conversives): इनका संबंध विनिमयता का होता है, जैसे: देना-लेना।
(ग) समानार्थकता/पर्यायता (Synonymy): जब एक अथवा एक जैसी वस्तु, पदार्थ या भाव के लिए एक से अधिक शब्दों का प्रयोग हो तो उनके बीच समानार्थकता या पर्यायता का संबंध होता है।

(घ) अनेकार्थकता (Polysemy): एक शब्द के एक से अधिक अर्थ होने की अवसथा अनेकार्थकता है।

(ङ) संकेतत्व और संपृक्तत्व (Denotation and Connotation): जब शब्द और अर्थ के बीच संकेतक और संकेतित का संबंध होता है तो इसे ‘संकेतत्व’ (अभिधेयार्थ) कहते हैं। किंतु जब शब्द और अर्थ का संबंध सीधा न होकर भाव या गुण के आधार पर निकाला जाता है तो इसे संपृक्तत्व (लक्षणार्थ) कहते हैं।

2.4 जान निरूपण : अर्थ-न्यास्थापन की तकनीकी विधि

आपूर्तिक युग में प्राकृतिक भाषा संसाधन और कृत्रिम बुद्धि के क्षेत्र में निरंतर अर्थ को न्यास्थित करने का प्रयास किया जा रहा है जिसे 1960 के दशक से अब तक देखा जा सकता है। साथ ही इसे निरूपित करने का भी प्रयास किया गया है जिसे ‘जान निरूपण’ नाम दिया गया है। जान निरूपण कृत्रिम बुद्धि के अंतर्गत आने वाली केंद्रीय संकल्पनाओं में से एक है। इसका उद्देश्य जान को न्यास्थित रूप से मशीन में संग्रहीत करना, उसका उपयोग करना और जान को इस तरह से निरूपित करना है कि उसके आधार पर तात्कालिक निरूपण लिए जा सके।

जान निरूपण बाह्य संसार की इकाइयों एवं प्रक्रियाओं को संकल्पनात्मक रूप से न्यास्थित एवं संरचित करने की एक विधि है। इसमें सम्मिलित किए जाने वाले जान के विविध प्रकार और उनकी निरूपण विधियाँ इस प्रकार हैं:

(क) वर्ग और वस्तुएं (Categories and Objects): वास्तविक संसार (world) की वस्तुओं और तथ्य (facts) को वस्तु (object) कहा गया है। जान निरूपण में वर्गों को वर्गीकरण अथवा अंतर्गत (taxonomic hierarchy) में संगठित या न्यास्थित किया जाता है। इसमें प्रत्येक वर्ग के छोटे-छोटे उपवर्ग (subcategories) होते हैं। फिर उस उपवर्ग के अपने छोटे-छोटे उपवर्ग होते हैं। किसी वर्ग की विशेषता उसके उपवर्ग और सदस्यों में भी पाई जाती है, जैसे : प्राणी एक बड़ा वर्ग है। मानव, पशु, पक्षी आदि इसके उपवर्ग हैं। ‘कौआ’ पक्षी उपवर्ग का एक छोटा उपवर्ग है। अब यदि प्राणी वर्ग की सभी वस्तुओं के संदर्भ में यह कहा जाए कि ‘प्राणी जलमें नहीं रहते हैं’ तो प्रत्येक कौआ संदर्भ में भी यह बात सही है, क्योंकि ‘कौआ’ प्राणी वर्ग के अंतर्गत आता है। किसी वर्ग की विशेषताओं को उसके उपवर्ग की वस्तुओं द्वारा ग्रहण करने की प्रक्रिया को अन्तर्वर्तन या विशालगति (inheritance) कहते हैं।
(ख) वस्तु और भाग (Object and Part) : भौतिक वस्तुओं में वस्तु और भाग संबंध भी पाया जाता है जिसमें एक वस्तु दूसरी वस्तु का अंग/भाग होती है, जैसे : ‘हाथ’ शरीर का भाग है, ‘अंगुल’ हाथ का भाग है। इसी प्रकार ‘भारत’ एशिया का भाग है और ‘उत्तर प्रदेश’ भारत का भाग है। ज्ञान निरूपण में इन संबंधों को भी निरूपित किया जाता है।

(ग) स्थितियाँ और घटनाएँ (Situations and Events) : कृत्रिम बूढ़ि के विकास के लिए स्थितियाँ और घटनाओं के सह-संबंधों का ढील-ढील निरूपण आवश्यक होता है। ज्ञान आधारित एजेंट (knowledge based agent) के लिए यह ज्ञान आवश्यक होता है कि स्थिति ‘क’ में ‘ख’ कार्य किया जाता है। वाणी संसार में होने वाले प्रत्येक कार्य एक ‘घटना’ है, जैसे : क्रिकेट खेलना, गिटार बजाना आदि। ज्ञान निरूपण में इनसे संबंधित ज्ञान को भी निरूपित किया जाता है।

(घ) मानसिक घटनाएँ और वस्तुएँ (Mental Events and Objects) : प्रत्येक व्यक्ति के मन में वस्तुओं और घटनाओं को लेकर कुछ प्रतिरूप (model) बने रहते हैं। ये प्रतिरूप उन्हीं घटनाओं, वस्तुओं और प्रक्रियाओं के बने होते हैं जो उस व्यक्ति के जीवन में घटित हुए हों। यदि किसी को कोई नई वस्तु दे दी जाए तो उसे उस वस्तु और उससे संबंधित प्रक्रिया के बारे में दूसरे से पूछना पड़ता है। अतः ज्ञान निरूपण में इससे संबंधित प्रतिरूप की आवश्यकता होती है।

(ड) अधिज्ञान (Meta-knowledge) : ज्ञान निरूपण में अधिज्ञान का अर्थ है- हम जो जानते हैं उसके बारे में जान।’

इस प्रकार ज्ञान निरूपण एक ऐसी विधि है जो अर्थ के प्रतिरूप को एक व्यवस्था प्रदान करती है। इसके माध्यम से आर्थिक विश्लेषण का कार्य स्वचालित रूप से किया जा सके यदि इसका लक्ष्य है। वास्तव में देखा जाए तो ज्ञान निरूपण अपने व्यापक परिप्रेक्ष्य में संपूर्ण ब्रह्मांड के ज्ञान को भाषिक इकाइयों के माध्यम से संरचित करके निरूपित करने की एक विधि है। इसके आधार पर मशीनी प्रणालियों द्वारा तात्कर्षित निर्देश दिए जा सकते हैं। जो सांसारिक ज्ञान हमारे मस्तिष्क में निहित होता है उसमें वस्तुओं के चित्र आदि भी होते हैं। ज्ञान निरूपण में निरूपित ज्ञान में ऐसी व्यवस्था नहीं है। इसमें समस्त ज्ञान के संरचित होने का आधार भाषिक इकाइयों के परस्पर संबंधों की व्यवस्था है।
2.5 ज्ञान निरूपण/अर्थ व्यवस्थापन की कुछ विधियाँ

ज्ञान निरूपण की कुछ प्रमुख विधियाँ को इस प्रकार से देखा जा सकता है:

(क) आर्थिक संजाल (Semantic Network)

यह संकल्पनाओं के बीच संबंधों को व्यक्त या निरूपित करने वाला एक संजाल है जिसे ‘साँचा संजाल’ (frame network) भी कहा जाता है। अँग्रेज़ी में इसे दो रूपों में लिखा जाता है - semantic net और semantic network. यह मुख्यतः प्रतिज्ञापर सूचनाओं के संबंधों में ज्ञान प्रतिरूपण की एक तकनीक है। इसी कारण कुछ विद्वानों द्वारा इसे प्रतिज्ञापर संजाल भी कहा जाता है। इनके माध्यम से शब्दार्थों को आर्थिक संबंधों के आधार पर एक दूसरे से सह-संबंधित किया जाता है और इस प्रकार से एक बड़ी संरचित इकाई का निर्माण किया जाता है।

यदि इसे ग्राफ के माध्यम से प्रदर्शित करे तो यह विभिन्न नोडों और उन्हें परस्पर सह-संबंधित करने वाले लिंकों के माध्यम से ज्ञान को निरूपित करने की एक तकनीक के रूप में प्रास होता है। इसमें नोड वस्तुओं (objects) को प्रदर्शित करते हैं और लिंक एकाधिक नोडों के बीच संबंधों को। प्रत्येक लिंक दिशा-निर्देशित होता है अर्थात उसकी एक दिशा होती है। साथ ही प्रत्येक लिंक को एक लेबल प्रदान किया जाता है। अत: आर्थिक ज्ञान को एक दिशा निर्देशित ग्राफ के रूप में समझा जा सकता है। सामान्यतः चित्रण करते हुए इन्हें समझाने में नोडों को गोले या बॉक्स द्वारा प्रदर्शित किया जाता है और लिंक इनके बीच रेखाओं को खींचकर बनाए जाते हैं। चूंकि आर्थिक संजालों में सभी नोड एक दूसरे से मिले हुए होते हैं अत: इन्हें सहचारी संजाल (Associative nets) भी कहा जाता है।

आर्थिक संजाल ज्ञान को संरचित करते हुए निरूपित करने की एक तकनीक है। इसके माध्यम से निरूपित ज्ञान में जो नोड और लिंक होते हैं वे सूचनाओं को आरेखीय रूप में प्रस्तुत करते हैं तथा दो संकल्पनाओं के बीच ताकिक संबंध स्थापित करते हैं। एक ग्राफ के माध्यम से हम सूचनाओं को कैसे प्रदर्शित कर सकते हैं, इसे Matthew Huntbach के एक ग्राफ द्वारा देखा जा सकता है।
इस ग्राफ में निम्नलिखित सूचनाएँ संचित हैं –

- Tom is a cat.
- Tom caught a bird.
- Tom is owned by John.
- Tom is ginger in colour.
- Cats like cream.
- The cat sat on the mat.
- A cat is a mammal.
- A bird is an animal.
- All mammals are animals.
- Mammals have fur.

ज्ञान के प्रतिरूपण की इस विधि को मानव भौतिक में निरूपित ज्ञान की विधि के अधिक करीब माना गया है।
(ख) वर्डनेट (WordNet)

वर्डनेट ऑँग्रेजी के शब्दों का एक विशाल डाटाबेस है। जार्ज ए. मिलर (George a. miller) द्वारा 1980 के दशक में इस परियोजना का आरंभ किया गया। इसमें शब्दों को उनके वर्गों, यथा- संज्ञा, विशेषण, क्रिया, क्रियाविशेषण आदि में समृद्ध करके रखा गया है जिसे संज्ञानात्मक पर्यायों का समूचन (सिनसेट – Synset) कहा जाता है। प्रत्येक सिनसेट एक प्रकार की संकल्पना का प्रतिनिधित्व करता है। सभी सिनसेटों को संकल्पनात्मक (conceptual-semantic) और शब्दव्यूहिक (lexical) संबंधों द्वारा एक दूसरे से जोड़ा जाता है। वर्डनेट ऑनलाइन प्रयोग के लिए इंटरनेट पर लिंक उपलब्ध है। इसमें शब्दों एवं सूचनाओं को देखने के लिए एक अंतरापृष्ठ भी दिया गया है –

[Image of WordNet search interface]

इस अंतरापृष्ठ में कोई भी शब्द या पदवंश इनपुट के रूप में दिया जा सकता है। उसके बाद आप जो भी सूचनाएं देखना चाहते हैं उन्हें “Display Options” में चयनित करें, जैसे यहाँ पर ‘Show all’ हाईलाइट है। इसमें सब कुछ प्रदर्शित किया जाएगा। उदाहरण के लिए नीचे ‘science’ शब्द देने के बाद प्राप्त परिणामों को दिखाया जा रहा है –
इसमें लिंक S द्वारा सिनसेट या आर्थी संबंधों को प्रदर्शित किया जा रहा है। प्रथम S को क्लिक करने पर आने वाली सूचनाइं स्क्रीनाइं इस प्रकार हैं:

<table>
<thead>
<tr>
<th>Sense</th>
<th>Word</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>S: (n)</td>
<td>science</td>
<td>(a particular branch of scientific knowledge)</td>
<td><em>the science of genetics</em></td>
</tr>
<tr>
<td></td>
<td>scientific discipline</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- *direct hyponym* / *full hyponym*
- *part meronym*
- *domain term category*
- *direct hypernym* / *inherited hypernym* / *sister term*
- *part holonym*
- *derivationally related form*

- *S: (n) skill, science* (ability to produce solutions in some problem domain) | *the skill of a well-trained boxer*, *the sweet science of pugilism* |
इसमें science शब्द के अर्थ के प्रयोग एवं विस्तार से जुड़ी अनेक प्रकार की सूचनाएँ दी गई हैं जिन्हें संबंधित लिंक पर क्लिक करके प्राप्त किया जा सकता है।

इस प्रकार देखा जाए तो वर्डनेट संकल्पनाओं एवं उनके बीच निहित संबंधों को व्यवस्थित रूप से प्रदर्शित करने वाली एक इकाई है। इसमें प्रत्येक शब्द से जुडी आर्थिक सूचनाओं की ऊपर दिखाए गए लिंकों के माध्यम से प्राप्त किया जा सकता है। अतः यह संगणकीय आर्थिक विश्लेषण के संबंध में एक उपयोगी दूर्देश है।

(ग) फ्रेम (Frame)

प्रणालियों के संदर्भ में फ्रेम की वात सर्वप्रथम मार्विन मिन्स्की (Marvin Minsky-1975) द्वारा की गई। वैसे इसकी संकल्पना संबंधी उनका प्रस्ताव उनके 1974 के "A Framework for Representing Knowledge." नामक आलेख में देखा जा सकता है। मिन्स्की के शब्दों में “a frame is a static data structure used to represent well-understood, stereotyped situations”। फ्रेमों के माध्यम से हम स्वयं के ज्ञान के संरचित रूप की कल्पना करते हैं। हम पुरानी परिस्थितियों के आधार पर अध्यायक आते तो वाली नई स्थितियों में भी अपने आप को एडजस्ट कर लेते हैं। नई परिस्थितियों में आई नई बातों को समझने के लिए हम पुराने अनुभवों की बातों का विस्तार से विश्लेषण करते हैं। इसी के आधार पर फ्रेम के निर्माण एवं प्रयोग की संकल्पना का विकास हुआ है।

साधारण शब्दों में फ्रेम एक प्रकार की डाटा संरचना है जिसका प्रयोग ज्ञान को विभिन्न उपसंरचनाओं बॉटकर इस प्रकार व्यवस्थित करने के लिए किया जाता है कि stereotyped situations को निरूपित किया जा सके। अतः यह 'ज्ञान निरूपण और तकनीक' की विधियों का ही एक भाग है। मूलतः फ्रेमों का विकास आर्थिक संजालों से ही किया गया है अतः इस्तेमाल प्रतिरूपण पर आधारित संरचनाएँ ही मानने अधिक उपयुक्त हैं। Russell और Norvig (2008) ने भी “Artificial Intelligence, A Modern Approach में कहा है, “structural representations assemble "...facts about particular object and even types and arrange the types into a large taxonomic hierarchy analogous to a biological taxonomy.”

फ्रेम की संरचना

एक फ्रेम की सामान्य संरचना इस प्रकार होती है:
ट्यापक रूप से देखें तो फ्रैम में सूचनाओं को संरचित रूप से प्रस्तुत किया जाता है। ये सूचनाएं कई प्रकार की होती हैं जो प्रत्येक फ्रैम के साथ जुड़ी होती हैं, जैसे- कुछ सूचना इसकी भी होती है कि फ्रैम का प्रयोग कैसे किया जाए। इसी प्रकार यह भी सूचना होती है कि आगे क्या घटित हो सकता है अथवा यदि अभिव्यक्ति सुनिश्चित ना हो तो क्या किया जाए। इस प्रकार की सूचनाओं को व्यवस्थित करने के लिए बनने वाली फ्रैम की संरचना नोडों और उनके बीच संबंधों के संजाल के रूप में बन जाती है।

(घ) स्क्रिप्ट (Script)

स्क्रिप्ट भी फ्रैम की तरह एक अवधारणात्मक इकाई है। इसका प्रतिपादन Schank और Abelson (1977) द्वारा किया गया। इसकी भी संरचना फ्रैम की तरह ही होती है। इसकी परिभाषा देते हुए कहा गया है, “A script is a structured representation describing a stereotyped sequence of events in a particular context.”

स्क्रिप्ट का प्रयोग प्राकृतिक भाषा बोधन (NLU) प्रणालियों में किया जाता है। इनके माध्यम से ज्ञान आधार को स्थितियों (situations) के ऐसे संचय के रूप में संगठित किया जाता है कि प्रणाली उनका बोधन कर सके अर्थात् समझ सके। यह एक प्रकार से प्रस्तूत लिंक्ड वाक्यों को शृंखला के रूप में संचित करने वाली इकाइयों हैं। इनमें किसी घटना के टाइप (type of occurrence) के लिए घटनाक्रमों (events) के क्रम को संग्रहीत किया जाता है। किसी स्क्रिप्ट के प्रत्येक घटनाक्रम को conceptual primitives के रूप में अभिव्यक्ति किया जा सकता है।

एक स्क्रिप्ट में निम्नलिखित घटक होते हैं:

- **प्रवेश शर्टें** (Prolog): जिनका स्क्रिप्ट को खोल करने के लिए सत्य होना आवश्यक है।
- **प्रॉप (Props)**: ये वस्तुएं (things) जो स्क्रिप्ट का कथय निम्नित करती हैं। अर्थात् स्क्रिप्ट में काम में आने वाली वस्तुएं।
- **रूपकारें** (Roles): अर्थात् ये कार्य जो प्रत्येक सहभागी द्वारा संपन्न किए जाते हैं।
- **दृश्य (Scenes)**: जो स्क्रिप्ट के वर्तमान पक्षों को प्रस्तुत करते हैं।
- घटनाएँ (events)
- परिणाम या तत्व जो स्क्रिप्ट के समापन पर ‘सत्य’ हो जाते हैं।

प्रत्येक दशक में कार्यक्रमों द्वारा कार्य किए जाते हैं। ये कार्य प्रोप की सहायता से किए जाते हैं।

अतः सभी को समेटिए। इनके पर स्क्रिप्ट को घटनाओं पर आधारित स्थितियों के संजाल या वृक्ष (network or tree of states) के रूप में निरूपित किया जा सकता है।

उदाहरण: रेस्टोरेंट का एक प्रसिद्ध उदाहरण:

<table>
<thead>
<tr>
<th>The classic example is the restaurant script:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scene: A restaurant with an entrance and tables.</td>
</tr>
<tr>
<td>Actors: The diners, servers, chef and Maitre d'Hotel.</td>
</tr>
<tr>
<td>Props: The table setting, menu, table, chair.</td>
</tr>
<tr>
<td>Acts: Entry, Seating, Ordering a meal, Serving a meal, Eating the meal, requesting the check, paying, leaving.</td>
</tr>
</tbody>
</table>

Schema प्रणालियों के आधार आधार, स्क्रिप्टों तथा आर्थिक संजाल का प्रयोग schema के संजाल के रूप में किया जाता है। सामान्यतः schema द्वारा भूमिकाओं को संपन्न करने वाली इकाइयों के बीच संबंधों को प्रदर्शित किया जाता है।

2.6 ज्ञान निरूपण और आर्थिक विशेषण

ज्ञान निरूपण का आर्थिक विशेषण की दृष्टि से अन्यतम स्थान है, विशेष रूप से तब जब यह कार्य स्वयंसेवक रूप से अर्थात् मशीन द्वारा किया जाना हो। यह मशीन में वस्तुओं एवं संकल्पनाओं को समझने, उनके बीच संबंध स्थापित करने तथा उनके संदर्भ में तकिक निर्णय लेने में आधारेभरत भूमिका का निर्वाचन करता है। अर्थ संबंधी किसी भी कार्य को यदि संगणकीय दृष्टि से देखा जाना हो तो ऐसी स्थिति में ज्ञान निरूपण की किसी न किसी इकाई की आवश्यकता पड़ती है। जैसा कि उपर ज्ञान निरूपण से संबंधित कुछ प्रमुख विधियों की चरचा की गई है। सभी में एक विशेष प्रकार की समानता देखी जा सकती है। वह है- ‘वर्ग-सदस्य’ निरूपित करने तथा वर्ग की विशेषताओं का सदस्य द्वारा ग्रहण किए जाने की विशेषता।

आर्थिक ज्ञान को वर्गों, उपवर्गों एवं सदस्यों आदि में विभक्त करने की विधि ज्ञान को संरचित करने की एक उपयुक्त विधि है। यह विधि न केवल मानव मस्तिष्क में संरचित संकल्पनात्मक ज्ञान के निकट है बल्कि प्राकृतिक भाषा संसाधन एवं तर्क की दृष्टि से भी उपयुक्त है। भाषिक इकाइयों को समझते हुए उनके बीच प्रासंगिक संबंधों को भी इसमें उचित स्थान प्राप्त होता है जिससे
कि उनके संबंध में तर्क करना अत्यंत सरल हो जाता है। अतः शब्द से बड़ी इकाई, जैसे- पदबंध, उपवाक्य या वाक्य आदि में आए घटकों एवं उनके बीच आर्थिक संबंधों के निर्धारण में इसकी केंद्रीय भूमिका हो सकती है। इसे ही ध्यान में रखते हुए प्रस्तुत शोध प्रवंश में इस प्रकार की एक संरचनात्मक इकाई का निर्माण किया गया है जिसे ‘जानआधार’ नाम दिया गया है। इस इकाई का प्रस्तुत विवेचन अध्याय 4 में देखा जा सकता है। इसके आधार पर ही संबंधवाची रचनाओं के विश्लेषण से संबंधित कार्य अध्याय 6 में किया जाएगा।

2.7 आर्थिक विश्लेषण: अन्य पक्ष

आर्थिक विश्लेषण भाषिक अध्ययन विश्लेषण एक प्रस्तुत पक्ष है। आधुनिक परिप्रेक्ष्य में प्राकृतिक भाषा संसाधन तथा कृत्रिम बुद्धि जैसे क्षेत्रों में इसका स्वरूप और प्रस्तुत हुआ है। ऐसी स्थिति आर्थिक विश्लेषण के संबंध में अन्य तकनीकों की विकसित हुई हैं। इनमें से दो प्रमुख इस प्रकार हैं:

(क) सत्वशास्त्रीय अर्थविज्ञान (Ontological semantics): यह आर्थिक विश्लेषण एवं अर्थ आधारित ज्ञान निरूपण की एक अत्यंत महत्त्वपूर्ण विधि है। ‘सत्वशास्त्र’ (Ontology) अपने आप में एक बहुत बड़ा क्षेत्र है और आर्थिक विश्लेषण तथा संगणकीय अनुप्रयोग से जुड़ा हुआ है। इस संबंध में ‘Sergei Nirenburg और Victor Raskin’ (2010) द्वारा कहा गया है:
“Ontological semantics is a theory of meaning in natural language and an approach to natural language processing (NLP) which uses a constructed world model, or ontology, as the central resource for extracting and representing meaning of natural language texts, reasoning about knowledge derived from texts as well as generating natural language texts based on representations of their meaning.”

इन विद्वानों ने इसकी संरचना और अनुप्रयोगात्मक पक्ष पर विचार करते हुए कहा है:

. The architecture of an archetypal implementation of ontological semantics comprises, at the most coarse-grain level of description:
• A set of static knowledge sources, namely, an ontology, a fact database, a lexicon connecting an ontology with a natural language and an onomasticon, a lexicon of names (one lexicon and one onomasticon are needed for each language);
• Knowledge representation languages for specifying meaning structures, ontologies and lexicons; and
• A set of processing modules, at the least, a semantic analyzer and a semantic text generator.
अतः सत्त्वशास्त्रीय विचार भी अपने आप में आर्थिक विश्लेषण एवं संगणकीय अनुप्रयोग का एक पक्ष है। इसकी कुछ बातें शोधकार्य में उपयोगी होंगी।

(ख) प्रच्छण्न आर्थिक विश्लेषण (Latent Semantic Analysis: LSA): इसका विकास Scott deerwester एवं उनके सहयोगियों द्वारा किया गया है। यह प्राकृतिक भाषा संसाधन के विभिन्न प्रलेखों (documents) के समुच्चय के बीच संबंधों के विश्लेषण और प्रलेख तथा terms से संबंधित संक्लपनाओं के समुच्चय के माध्यम से उनमें निहित terms के उद्धार की एक तकनीक है। इसका प्रयोग मुख्यतः ‘वेक्टर आधारित अर्थविज्ञान’ (Vectorial Semantics) में किया जाता है। इस विधि में हम यह मानकर चलते हैं कि अर्थ की दृष्टि से समान शब्द समान प्रकार के पाठ खड़ों में प्राप्त होते हैं। अतः उनसे संबंधित matrix का निर्माण किया जाता है जिसमें प्रत्येक अनुच्छेद में आए शब्दों की संख्या को रखते हैं। इसमें रो तथा कॉलम का निर्माण करते हैं। रो में अद्वितीय शब्दों (unique words) और कॉलम में अनुच्छेदों को रखा जाता है जो किसी बड़े पाठ से लिए जाते हैं। इसके लिए एक गणितीय विधि का प्रयोग किया जाता है जिसे singular value decomposition (SVD) कहा जाता है।

LSA मुख्यतः पाठों की दिशात्मक आधार पर तुलना, प्रति भाषा प्रत्यय (Cross Language Retrival), शब्दों के बीच संबंध स्थापना (जैसे: पर्यायता एवं बहुअर्थकता) एवं सूचना प्रत्यय आदि में बहुत अधिक उपयोगी है। इसके द्वारा शब्दों के समूहों के बीच सामानताओं को अर्थ की दृष्टि से बताने का अत्यंत सरलता से किया जाता है। इस कारण MCQ प्रश्न उत्तर मॉडलों में बहुविकल्पीय प्रश्नों को समझने में भी यह उपयोगी है। आर्थिक मॉडलिंग की दृष्टि से इसके उपयोगी पक्षों को देखे तो इसका प्रयोग मानव स्मृति के अध्ययन में भी उपयोग चिह्नित कर सकता है।

इसके अलावा आर्थिक विश्लेषण की दृष्टि से कुछ और भी कार्य हुए हैं जिनमें ‘व्यक्त/आर्थी विश्लेषण’ (Explicit Semantic Analyasis), ‘प्रमुख घटक विश्लेषण’ (Principal Components Analysis) संभाव्यतात्मक प्रच्छण्न आर्थिक विश्लेषण (Probabilistic Latent Semantic Analysis) आदि प्रमुख हैं। आगे Malka Rappaport Hovav (2010) आदि ने ‘घटनात्मक अर्थविज्ञान’ (Event Semantics) की भी बात की है। इसी प्रकार Rochelle Lieber (2004) द्वारा ‘सूचनापरक अर्थविज्ञान’ (Informational Semantics) का उल्लेख किया गया है।
2.8 उपसंहार

इस प्रकार प्रस्तुत अध्याय में आर्थी विश्लेषण के विविध पक्षों को प्रस्तुत किया गया। सर्वप्रथम हमने इसमें यह देखने का प्रयास किया कि अर्थ क्या है और मानव मस्तिष्क में इसकी क्या स्थिति है। इस संदर्भ में हमने इसे एक मानसिक संकल्पना के रूप में देखा। यह संकल्पना ‘बिंब’ के रूप में मानव मस्तिष्क में निहित रहती है। मस्तिष्क में इसके स्थापित होने की प्रक्रिया ‘बिंबनिर्माण’ है तथा इसकी सहायता से किसी भाषिक उक्ति के अर्थ का निर्धारण ‘अर्थबोध’ है। मानव मस्तिष्क में आर्थी संकल्पनाएँ न केवल अंकत होती हैं बल्कि एक दूसरे से सहसंबंधित होकर संरचित होती हैं। इनके आधार पर संकल्पनाओं के बीच मिलन-मिलन प्रकार के संबंध निर्मित होते हैं, यथा- अवनामिता, विलोमता, पर्यायता वा अंगांग संबंध आदि।

अर्थ मस्तिष्क में संरचित रूप से संचित है। इसमें कोई संदेह नहीं। किंतु इस ज्ञान को मशीनी अनुप्रयोग के दृष्टिकोण से संरचित करने की बात की जाए तो इसके लिए जो प्रविधि प्रयोग में लाई जा रही है उसे ‘ज्ञान निरूपण’ नाम दिया जा रहा है। इसके अंतर्गत कुछ प्रसिद्ध विधियों में आर्थी संजाल, वर्डनेट, फ्रेम और स्क्रिप्ट आदि हैं। इसके अलावा आर्थी विश्लेषण हेतु किए गए प्रयासों को देखा जाए तो ‘प्रचलित आर्थी विश्लेषण’ अथवा ‘वेक्टर आधारित अर्थविज्ञान’ (Vectirial Semantics) आदि कुछ विधियों प्राप्त होती हैं। प्रस्तुत शोधकार्य में ज्ञान निरूपण आधारित तकनीकी का प्रयोग किया जाएगा।