Chapter II

DOMAIN ANALYSIS
AND METHODOLOGY

2.1 Domain based MT: English-Hindi in the Share-market

2.1.1 Need for domain-based MTSs

Building a Machine Translation System (MTS) for register specific translation has been one of the prime motives of the MT researchers. Most of the early MT systems like GAT, CETA, TAUM were meant for translating technical texts. As most commercial applications of MTS often relate to the technical areas like scientific texts, meteorological data, weather forecasts, defence, environmental, share-market & business and other similar domains (refer to section 1.2.4). Designing MTSs for such domains is easier and cost-effective with the product finding a ready market.

The present research for designing a Lexical Semantic Component (LSC) for English-Hindi translation in the domain of share market is
an effort on similar lines. The share-market and business language in India is primarily English, while the majority speaks and understands Hindi. Thus to link this very important part of our daily activity to the Hindi speaking masses, the system-design proposed would be worth undertaking. The advantage of designing the proposed domain specific LSC are the following -

-the LSC can be interfaced with an English parser and a Hindi generator to directly translate the share-market register of English into Hindi,

-the LSC can also be used to grant domain-competence to any general purpose interlingual MTS, and

-the LSC can be easily augmented and extended to include more languages.

2.1.2 English-Hindi MTS for Share-market

Share-market and business represent a well specified domain of linguistic usage. These registers of English and Hindi present technical terms which have domain specific meaning, for example, ‘sensex’, ‘deben-
ture', 'foreign institutional investors', 'stock-exchange', 'share-market',
'bourse', 'stag', 'exim', 'operator', 'scrip' etc. (refer to appendix for a
detailed listing).

The register is full of metaphorical usages like 'bullish trend in share-
market' (a situation of continuous rise in share-prices), 'chAndI uchhallI'
(spurt in silver prices) whose literal meanings may interfere with the real
import. Besides, there are some special sentences and phrases used for
communicating something in a particular context. Handling this specialized,
technical variety of the languages in question for MT purpose was a
problem. Therefore to make any general purpose MTS effective in this
domain, it was necessary to grant domain-competence in the form of a
component which would represent lexical items from the domain in an
abstract manner along with provisions for syntactic information. The
proposed LSC is an effort on similar lines.

2.1.3 **LSC for English-Hindi MTS for Share-market**

A Lexical Semantic Component (LSC) for English-Hindi Machine
Translation in the share-market domain has been formalized using LCS
representation technique (implementation to follow later). The LSC has
conceptual representations for typical lexical items from the specified
domain. In particular, the entries will be of verbs in the form of Root
LCSs (or RLCSs) with a compositional/decompositional algorithm to derive Composed LCSs (or CLCSs) from the RLCSs or vice versa. This LCS-based LSC for typical share-market register of Hindi and English can also be interfaced with a general purpose LCS-based interlingual Machine Translation System to make it translate effectively in the specific domain of share-market & business (refer to section 1.3.1). Since language is universally acknowledged as the most dynamic of human behaviours, and the view that translation is more of an artistic enterprise than mechanical, the research has taken a central position in proposing a componential approach to the problem - that there should be small components of domain specific (preferably of the technical types) LSCs to grant domain competence to a general purpose MTS, and/or to undertake domain specific translation.

2.2 Domain Analysis and Divergence

2.2.1 Methodology followed

To begin with, the domain analysis was done by carefully studying the pattern of language use from a range of standard sources as listed below -
-Economic Times (daily) : April 1995-March 1996 issues,

-Business India (fortnightly) : April 1995-March 1996 issues,

-Business & Economy pages of HT, and TOI : Jan 1996 - March 1996

-Times FM’s "Invest Times", market news programme on All India Radio aired at 0750 hrs IST : March 1996

Glossaries were prepared for nouns (565) verbs (300) and adjectives (150). These glossaries representing the register of share-market & business were then compared with the following standard dictionaries -

- OUP concise dictionary of business world

- Dictionary of economics & commerce (by Hanson J.L., London, 1971)

- Dictionary of business & management (by Parikh K.C., Bombay, 1972)

- Dictionary of international economics (by Siegfried K., Leiden, 1976)
As a next step, four students related to economics, business and share-market were told to read the current literature on business & markets with reference to these glossaries and see if there are any new usages.

The next stage was searching for lexical equivalence in Hindi. This was another problem in itself as optional terms are often used for Hindi in different sources, particularly the media and the available bilingual dictionaries of business & economics. For ease of communicability, it was decided to use the more popular lexical options. For example, the Hindi equivalents for ‘share’ are ‘pratibhUti’ and ‘share’ itself. Since the latter is more popularly used, its choice as the Hindi equivalent was preferred.

The sources (bilingual dictionaries & glossaries) used for collecting Hindi terms from business and market are as follows:
After data sampling and fixing the lexical equivalences, the conceptual representations for them were formalized using the technique of Lexical Conceptual Structure as formulated by Jackendoff and later updated by
Dorr1. The entries called the RLCS (Root Lexical Conceptual Structure) were arranged according to the syntactic category. Thereafter, a to-and-fro derivational mechanism was provided for to account for back and forth composition and decomposition of the conceptual structures. Each RLCS was also provided the syntactic information useful for mapping form the LSC to the syntactic structure of the target language.

2.2.2 Typicality of register

The share-market register of English is fundamentally of a technical type. There are many such vocabularies and phrases which typically form a large part of business & market dictionaries. Such words are largely used in this and related domains only. However, there are also a considerable number of words which may be from the language in general but have a specialized and unique meaning in the given domain.

The domain abounds in metaphorical usage of the language which poses problems for semantic representation and translation (ref section 3.1.5).

NOMINALS
******

Examples of specialized technical vocabulary from the given domain along with their Hindi equivalents

(* pl) (+ der) (pl - plural)
(adj - adjective) (h. - Hindi equivalent)

acid-test-ratio. h. saakh-nirdhaaraN-anupaat
blue-chip-share. * pl -s h. blu-chip-sheyar
bonus-share. * pl -s h. bonus-sheyar
bourse.* pl -s h. buurs
cash-share. * pl -s h. nakad-sheyar
convertible-debenture.* pl -s parivartaniiya-RN-patra
convertible-whole-life-assurance-policy.* pl -ies
cyclic-share. * pl -s h. chakriiya-sheyar
debenture. * pl -s h. RN-patra
equity.* pl -ies h. ikwiti
exim. h. ekzim
exim-policy. * pl -s h. ekzim-polic
foreign-institutional-investor.* pl -s
    h. videshii-sansthaanik-nivesh-kartaa
fully-convertible-debenture. * pl -s
    h. puurNa-parivartaniiya-RN-patra
group-share. * pl -s h. samuuha-sheyar
List of domain vocabularies with unique domain dependent meanings
alongwith their Hindi equivalents

(* pl) (+ der) (pl - plural)
(adj - adjective) (h. - Hindi equivalent)

bear+ adj -ish h. biyar
bond.* pl -s h. bandh-patra
bonus-candidate.* pl -s h. bonus-pratyasheem
capital.* pl-s h. puunjii
company.* pl -ies h. kampani
conversion-price.* pl -s h. parivartan-bhaav
hawala. h. hawaalaa
loss.* pl -es h. ghaaTaa
portfolio.* pl -s h. porTfolio
security.* pl -ies h. pratibhuuti
share.* pl -s h. sheyar
speculation.* pl -s h. sāTTaa
term-loan.* pl -s h. aavdhik-karz
territory.* pl -ies h. kshetra

VERBS

*****

The semantic range of arguments of the verbs gets delimited. For eg verb ‘acquire’ will have as argument only material things as asset, property etc or things denoting prosperity, well-being etc.

Examples of verbs with technical domain-dependent meaning along with their Hindi equivalents

vi(vt) -> the verb though intransitive, is however used as transitive
vt/vi -> mostly used as transitive, rarely as intransitive
vt, vi -> used in both conditions

budget. vi. h. bazaT-karnaā
buy. vt. h. khariid-naa
confiscate. vt. h. zabt-karna
export. vt. h. niryaat-karna
finance. vt. h. puunjii-lagaanaa
fund. vt. h. puunjii-dena
higgle. vt. h. mol-bhav-karna
hike. vt. h. baRhaanaa
hire. vt. h. bhaaRaa-karna
import. vt. h. aayaat-karna
incur (loss). vt. h. (ghaaTaa) sahnaa
invest. vt. h. puunjii-lagaanaa
offload. vt. h. bhaar-kam-karna
profit. vi. h. laabh-paanaa
roll-out. vi/vt. h. nirmit-karna
sell. vt. h. bech-naa
slash. vt. h. kam-karna
spend. vt. h. vyay-karna
spurt. vi. h. uchhal-naa
tax. vt. h. kar-lgaanaa
trade. vi. h. vyaapaar-karna
waive. vt. h. maaf-karna
Examples of verbs with a unique meaning in the given domain (in the sense that only a domain determined set of arguments will be selected for sentences) along with their Hindi equivalents

vi(vt) -> the verb though intransitive, is however used as transitive
vt/vi -> mostly used as transitive, rarely as intransitive
vt, vi -> used in both conditions

account. vt/vi . h. lekhaa-karan-karna
acquire. vt. h. arjan-karna
advance. vi/vt. h. agrim-denaa
borrow. vt. h. udhaar-lenaa
bounce. vi. h. lauT-naa
cost. vt. h. kiimat-lenaa
crash. vi. h. TuuT-naa
exchange. vt. h. vinimay-karna
fare (well). vi. h. (achhaa) karna
flare-up. vi. bhaRaknaa
grow. vt(vi). h. baRhnaa
hot-up. vi. h. garmaanaa
jump. vi. h. uchhalnaa
launch. vt. h. srigaNesh-karna

64
market. vt. h. bech-naa/baazaar-men-denaa
operate. vi. h. sanchaalan-karna
own. vt. h. svaamii-hona
pay. vt. h. bhugtaan-karna
quote. vt. h. koT-karna
record. vt. h. darz-karna
register. vt. h. darz-karna
register. vi. h. naamaankan-karna
relaunch. vt. h. punah-srigaNesh-karna
shine. vi. h. camaknaa
tie-up. vi. h. baandhnaa
unload. vt. h. halkaa-karna

2.2.3 Fixing lexical equivalences

As is the case with most rechnical domains of Hindi usage, this domain too has a lot of English words being used optionally. Their Hindi counterparts though available, are more difficult and cumbersome. Hence they are less preferred compared to English vocabularies. The following lists will clarify more -
List of English words which are optionally used as Hindi equivalents (partially or wholly)

NOMINALS

********

bear+ adj -ish h. mandaRiyaa. op. biyar
blue-chip-share.* pl -s h. blu-chip-sheyar
bond.* pl -s h. bandh-patra. op. bonD
bonus-candidate.* pl -s h. bonas-pratyashii. op. bonas-pratyashe
bonus-share.* pl -s h. bonas-sheyar
bourse.* pl -s h. buurs
budget.*+ pl -s +der-n -ing h. bazaT
bull+ adj -ish h. tejaRiyaa. op. bul
cash-share.* pl -s h. nakad-sheyar
certificate.* pl -s h. serTifiket
chamber.* pl -s h. prakoshTh. op. chamber
company.* pl -ies h. kampani
cyclic-share.* pl -s h. chakriiya-sheyar
discount.*+ pl -s n -ing h. baTTaa. op. DiskaunT
equity.* pl -ies h. ikwiti
equity-share.* pl -s h. ikwiti-sheyar
euro-issue.* pl -s h. yuuro-ishu
exim h. ekzim
exim-policy.* pl -s h. ekzim-polici
FII.* pl -s h. ef-ai-ai
fund.*+* pl -s n -ing -s h. puunjii/nidhi. op. fanD
gift-tax.* pl -es h. gifT-kar. op. gifT-Taks
group-share.* pl -s h. samuha-sheyar
growth-share.* pl -s h. vRddhi-sheyar. op. groth-sheyar
hawala h. hawaalaa
income-share.* pl -s h. aay-sheyar. op. inkam-sheyar
index.*+* pl -ices n -ing -s h. suuchii. op. indeks
indira-vikas-patra.* pl -s h. indiraa-vikaas-patra
issue.* pl -s h. ishu
jeevan-dhara-policy.* pl -ies h. jiivan-dhaaraa-polisi
jensen-measure.* pl -s h. zensen-maap
jobber.* pl -s h. zobar
jobbing.* pl -s h. zobing
kisan-vikas-patra.* pl -s h. kisaan-vikaas-patra
market.* pl -s h. baazaar. op. maarkeT
order.* pl -s h. aadesh. op. orDer
portfolio.* pl -s h. porTfolio
preference-share.* pl -s h. preferens-sheyar
premium.* pl -a h. primiam
private-sector.* pl -s h. nijii-kshetra. op. praivet-sektor
public-sector.* pl -s h. pablik-sektor
reserve.* pl -s h. rizerv
scrip.* pl -s h. skrip
SEBI h. sebi
sector.* pl -s h. kshetra. op. sektor
sensex h. senseks
share.* pl -s h. sheyar
share-market.* pl -s h. sheyar-baazar

VERBS AND ADJECTIVES

There are very few verbs and adjectives which are used optionally as their Hindi equivalents. For example -

VERBS

charge. vt. h. kiimat-lagaanaa/aarop-lgaanaa. op. chaarz-karna
chargesheet. vt. h. aaropit-karna. op. caarzshiiT-karna
control. vt. h. niantraN-karna. op. kanTrol-karna
export. vt. h. niryaat-karna. op. eksporT-karna
import. vt. h. aayaat-karna. op. imporT-karna
ADJECTIVES

**********

automatic. h. svatah. op. auTomaTik
bogus. h. farzii. op. bogas
multinational. h. bahuraaSTriy. op. malTinashanal
national. h. raaSTriya. op. nashanal
private. h. nijii. op. praiveT
public. h. saarv-janik. op. pablik
standard. h. maanak. op. sTanDarD
transnational. h. Transnashanal

While most Hindi terms were picked up from the primary sources from the Hindi business media, some had to be taken from official Govt of India terminology bank ("A Glossary of Commerce", CSTT, HRD, GOI publication). While these glossaries are mostly prepared by a team of qualified experts and researchers, there is a tendency to coin difficult and cumbersome terms in some cases instead of accepting more popular and simpler English terms. For this research, we have opted for more popular English terms which are being used in daily business language. For example, the terms for ‘bear’ is ‘mandaRiaa’ which is rarely found used even in the most ordinary variety of share-market language. Other examples are ‘bull’ (tejaRiaa), ‘launch’ ((jalaavtaraN) which is incorrect in this domain).
There are many terms in the official CSTT glossary of commerce which needed simplification. For example, the Hindi for the compound ‘fixed asset turnover ratio’ is given as ‘sthir parisampatti aavart anupaat’. Now in this and similar cases, the use of ‘pari-’ as prefix to ‘sampatti’ is not required as the word ‘sampatti’ also gives the same meaning as a constituent in the compound. Other examples are ‘quantum’ (pra-maatraa), ‘bulk’ (pra-punj/baRii maatraa).

Searching for Hindi equivalents from the business sections of Hindi newspapers/magazines was a curious exercise. The Hindi writings in this domain does not appear to be restricted by the paucity of technical terms. While on the one hand it liberally accepts English terms wherever necessary for simplicity and clarity of communication, it also uses many native terms some of which are highly metaphorical in nature. For example-

-moong uchhlii (price of moong (a pulse) rises)
-sonaa mandaa (dullness in gold-market / price of gold dips)
-masoor bhaDkii (price of masoor (a pulse) flares-up)
-chaandii chamkii (silver-market brightens/silver-prices brighten)
Since this research is a part of English to Hindi Machine Translation, the use of such loaded verbs have been incorporated as equivalents of suitable English counterparts. Since the domain of language use is very specified and limited in nature, we hope that this will take care of most of the technical language used in business and market literatures for translation into a Hindi which is currently being used in business sections of newspapers and magazines.

2.2.4 On Hindi syntax

Hindi, an MIA (Modern Indo Aryan) language has a sentential order S-O-V which is said to be partially free because of its inflectional retention from the Indo-Aryan heritage. This characteristic of the language is explained by the property that in a sentence, noun and adverbial phrases occur arbitrarily with the verb occurring at the end. The grammatical relations subject, object, indirect object, subject and object complements which are realized by noun phrases appear in arbitrary order in a sentence but the order of the constituents inside a phrase is almost fixed.

Among other features of Hindi syntax, agreement is perhaps the most problematic. Unlike English, the Hindi verb phrase sometimes agrees with the object instead of with the subject. The agreement in Hindi may take place either between immediate constituents or distant constituents.

At the phrase level, the agreement takes place between a modifier and a
noun it modifies. A modifier may be an adjective, noun in genitive or/and participle. An adjective agrees with the noun it modifies in number, gender and case. When there are more than one adjectives modifying a noun, then all the adjectives agree with the noun they modify in number, gender and case. When there are more than one nouns of variable gender, number and case, the adjective agrees with the first of these nouns it modifies. The genitive case marker agrees with the word following to which it is related in number, gender and case. In a verb phrase, aspects and tense markers agree in number, gender and person. The aspect and mood markers agree number and person.

At the level of sentence, distant constituents are related to one another in number, gender, person and case. This relationship between constituents can be classified as agreement between subject & verb, object & verb, pronoun & noun, adverb & subject, and adverb & object. In case of an unmarked subject noun phrase, the verb agrees with the subject. In case of a single noun denoting the subject, verb agrees with it in number, gender and person. But in presumptive future, verb does not agree with subject in gender. In case of more than one noun with same number, gender and person denoting the subject, the verb takes the plural form of the same gender. When there are two nouns differing in gender, the verb generally takes the masculine plural form. When there are several nouns in masculine and feminine denoting the subject, then the verb agrees with the nearest noun in number, gender and person. In case of two abstract nouns, the verb optionally takes singular form. When there is
noun or pronoun in apposition to the conjuncts, the verb agrees with the conjunct. If in a sentence, the noun phrase denoting subject is marked and noun phrase denoting object is unmarked, the verb agrees with the object. When there is a single noun denoting object, the verb agrees with the object in number, gender and person. When there are more than one noun denoting the object, the verb agrees with the object nearest to it. In case of a ditransitive verb, the verb agrees with the direct object. In case there are no unmarked noun phrases in a sentence, the verb takes the third person singular number and masculine gender. The pronoun agrees with the noun it refers to. Pronoun takes the first person when the subject says something about himself and it agrees in number with the subject. The pronoun takes second person when something is being said about the object and it agrees with the object in number. The adjectives as well as present and past participle used as adverb agrees with the subject in number in case the subject is unmarked. The adjectives as well as present and past participle used as adverb agrees with the object in number and gender in case the subject is marked and object is unmarked.  

2.2.5 Problems of translating

2.2.5.i Handling metaphorical usage

As seen in the previous section, the language of share-market (both English and Hindi) is highly metaphorical thus making translation more difficult. Let us look at some of the examples for both English and Hindi:

English

Sugar hardens
(chiinii pakkii-PRES)

Market turns bearish
(baazaar muR-PRES mandaa)

Shares react, sensex down 114.4
(sheyar-PL-NOM pratikriya-vyakt-PRES, senseks neeche 114.4)

Bears dominate the market
(mandaRiya-PL chhaa-PRES Det baazaar-par)

Shares remain in the bear’s grip
Forex reserve falls to $21.39 bn
(videshi-mudraa-bhanDaar giraa tak $21.39 arab)

PUS stocks recover on FIIs’ buying
(PUS-bhanDaar sudhar-PRES par FII-PL-GENT khariid)

US $ closed at 35.65
(amriikii-Dollar band-PAST par 35.65)

Bad deliveries hit the market
(kharaab khep-PL chot-kar-PRES Det baazaar)

moong ThanDhii
(muung cool)

masuur bhaRkii
(masuur flare-PAST)

sonaa chamkaa
(gold shine-PAST)

chaandii tez
(silver fast)

baazaar mandaa
(market dull)

tel naram
(cooking-oil soft)

chiinii TuuTii/sudhrii
(sugar break/improve-PAST)

guR garam
(gur hot)

aaluu mazbuut/kamzor rahaa
(potato strong/weak remain-TA)

taambaa sambhlaa
(copper stabilize-PAST)

sheyar luRhke
(share-PL crash-PAST)

suurajmukhi-tel 60 rupaye se luRhkaa
(sunflower-oil 60-rupee-PL by crash-PAST)

juuT men mande kaa jhaTkaa
(jute in dullness of shock)

sarrafaa ghaTaa
(jewelry-market decrease-PAST)

rupayaa naram
(rupee soft)

2.2.5.ii Linguistic divergences

Since Hindi and English belong to different families and cultural backgrounds, the divergences are imminent. Dorr(1993)\(^1\) lists seven types of divergences in translation while discussing the efficacy of LCS formalism in the UNITRAN model of machine translation at MIT. For the

\[ \text{---} \]

present research which operates in a strictly technical domain, we have found mostly syntactic, lexical and metaphorical divergences that pose problems for translation:

1. **Syntactic**

**Word-order:**

The S-V-O (English) and S-O-V (Hindi) has been a much dogged problem for translators. Associated with this are English pre- and Hindi post-positions.

E. poor economic growth have hit the share-market

\[
\begin{array}{ccc}
S & V & O \\
\end{array}
\]

H. kharaab aarthik-vikaas-ne sheyar-baazaar-ko

\[
\begin{array}{cccc}
S & \quad & O \\
\end{array}
\]

duSprabhaavit kiyaa hai

\[
\begin{array}{c}
\end{array}
\]
Categorial:

In terms of syntactic category mismatch, which occurs not so infrequently, the resulting sentence may be of an entirely different construction with consequent change in meaning.

As an example of syntactic category mismatch, we have verbs like 'cost' and 'own' which become nouns 'kiimat' and 'maalik' in Hindi at the cost of structure and meaning:

E. I own this company

H. mam is kampani kaa malik huuN
   I this company of owner am

In this case, the verb 'own' has as equivalent a noun in Hindi 'maalik' which causes all this problem and change of meaning in translation. The resultant sentence therefore is 'I am the owner of this company' in place of 'I own this company'
2. Lexical

There are many words in this domain which when translated in to Hindi have equivalents showing slight variations in meaning. For example -

inflational : sfiiti-sambandhi (‘inflation relating’)

risky : jokhim-bharaa (‘full of risks’)

3. Metaphoric

In most cases, the use of verbs and other lexical items is done in an indirect sense. Such usages have been long accepted as part of daily business parlance with no consequent mis-interpretations. The domain specific words shown in the following lists are loaded and are generally used in indirect and metaphorical senses, e.g.

NOUNS

****

bear+ adj -ish h. biyar/mandaRiyaa
blue-chip-share* pl -s h. blu-chip-sheyar
bull+ adj -ish h. bul/tejaRia
constant-mix-policy* pl -ies h. lagaataar-mishraN-niiti
conversion-price* pl -s h. parivartan-bhaav
liquidity* -ies h. dravtaa
open-order* pl -s h. khulaa-aadesh
operator* pl -s h. sanchaalak
portfolio* pl -s h. porTfolio
tax-treatment* pl -s h. kar-upchaar

ADJECTIVES
**************
aggressive
bearish
bullish
close-/open ended
vertical

VERBS
*****
bounce. vi h. uchhalkar-lauTnaa
close. vt h. band-honaa
crash. vi h. dharaashaayii-honaa
cripple. vt h. apaahiz-banaanaa
crop-up. vi h. aanaa
flare-up. vi. bhaRaknaa
float. vt. h. bahnaa
gear-up. vt. h. taiyaar-honaa
harden. vi. h. pakka-honaa
hit. vt. h. choT-karnaa/parnaa
hot-up. vi. h. garmaanaa
jump. vi. h. uchhalnaa
lock-horns. vi. h. jhagRaa-karnaa
offload. vt. h. utaarnaa
roll-out. vi/vt. h. banaanaa
shine. vi. h. camaknaa
shock. vt. jhaTkaa-lagnaa
unload. vt. h. bhar-utaarnaa