CHAPTER – 4
IDIOMS IN DISCOURSE

CONTENTS:

4.1. INTRODUCTION
4.2. MOTIVATION
4.3. CORPUS ANALYSIS
4.4. METHODOLOGY
4.5. EVALUATION
4.6. CONCLUSION
4.1. INTRODUCTION

In the earlier chapters, I have observed that the Idiomatic phrases are in general idiosyncratic. Hence, they need the relevance of some extra information from discourse usually occurring before or after the idiomatic phrases. The information beyond the probability of an idiom in order to identify and interpret it as an idiom may be termed as discourse. Therefore, to see this we intend to study the distribution of idioms in the corpus of actual texts and see how they are usually identified as idioms and the availability of the context.

It is often noted that idiomatic multi-word expressions are lexicalized and not always have a special status. Idioms traditionally involve lexical items of fixed metaphor as in “it is raining cats and dogs”, or an unusual formation, for example, “to dress up to the nines” (to dress in one’s best clothes). These expressions contain idioms and not that the entire text is an idiom. At times, the idiom is somewhat more motivated (i.e. transparent or easily interpreted) than its foreign counterpart that appears relatively opaque and arbitrary. For example, 1. “a fat lot of good that’ll do me” (meaning ironically and informally ‘that is completely useless’) is relatively predictable. According to traditional accounts, idioms resist changes in word choice or differ in the extent to which the expression can be transformed. Thus, because one cannot say, “? I am done a fat load of good by that”, or “? a great load of good that might do me”, example, (1) can be said to be idiomatic on lexical and syntactic grounds.

Lexical and semantic properties and a sense of uniqueness are not the only defining features of these expressions. Expressions, as implied by Fillmore et al. (1988) have a life of their own in the language, and can survive even when truncated or reformulated. While the two grammatically similar expressions 2. “I have had it” and 3. “I have done it”, have the same grammatical structure; only one of these is recognized as an expression as such. 2. “I have had it” can be taken to mean, “I have had enough”, and this intention cannot be translated literally into another language with the same effect. But the example 2. is also part of a longer expression, “I have had it up to here” (again, a long sequence of closed
class items), although native speakers do not have to access the whole expression to realize its rhetorical potential. Clearly, in order to have this vocabulary (although the longer version can exist in various truncated parts: “I have had it”, “I’ve had it”, and (accompanied by an appropriate gesture), “Up to here”). But this is not the most salient feature of the idiomatic expression. The main difference between 2. and 3. is that utterance 2. has a conventional rhetorical meaning. The move from word sequence (as a sentence) to rhetorical unit (as an utterance) is a central tenet of speech act theory, and certain linguists have claimed that this property, sometimes termed idiomaticity, is more central to a concept of native-like language use than the principles of grammatical competence proposed in mainstream linguistics (proponents include Yorio (1980), Pawley and Syder (1989), Sinclair (1991), Makkai (1992)). Sinclair’s definition of idiom, for example, posits that language is in constant flux (synchronically and diachronically), in a cycle between the compositional “open choice” of single words, to the automatic and “idiomatic” use of whole expressions.

While traditional accounts of idiom concentrate on semantic opacity, others have explored the role of idioms in discourse. Makkai (1972), for example, emphasized the distinction between lexemes with predictable semantics, such as “fly of the handle” and the expressions with some rhetorical force, such as “not a mouse stirred”. The correct interpretation of either 2. or 3. above, equally depends on the extent to which they obey the general Gricean principles of conversation (cf. Grice, ). As Moon (1992) points out, when utterances such as 2. or 3. appear to contravene the principles of relevance, the reader or interlocutor is justified in searching an alternative interpretation. This shift in emphasis has the advantage of making the concept of idiom less categorical. It means that sentence 2. is typically interpreted as “I have had enough” unless the literal sense “I have had it” may make sense in context (for example, as a response to 2a. Have you had your measles injection?). Utterance 3. I have done it, is typically interpreted as literal, if no relevant interpretation is forthcoming. The idea of authenticity and naturalness, as with “typicality” is a principle enshrined by empirical linguists such as Sinclair.
According to Moon (1992), idioms play a vital role in encoding modality not only as potential speech acts, but as alternative and marked formulations in a system of choices of expression. For example, “to walk slowly” can be encoded subjectively as “to walk at a snail’s pace”, where the use of the idiom can be interpreted as an additional subjective evaluation of the preposition. For Moon (1992), the paradigmatic choice of expression by an idiom as opposed to a more literal expression always implies some rhetorical force, and this explains the large number of idioms used as euphemisms or intensifying expressions (one thinks here of idioms for taboo subjects such as “death to shuffle off the mortal coil”, “emotional status to live it up” or “to bark up the wrong tree”). Moon (1992) and Fernando (1996) further classify idioms according to Halliday’s meta-functions expressions which convey (a) ideational information as in “down in the dumps”, (b) interpersonal or dialogic information as in “at the end of the day”. From this perspective, dictums, cliches and “turns of phrase” can be seen to be archetypal idioms.

Although both Moon (1992) and Fernando (1996) point out the rhetorical role of idioms, they nevertheless stick to the traditional criteria for inclusion into the category: syntactic and semantic uniqueness. It is equally valid to see rhetorical function and pragmatic force as determining factors in the classification of idioms, although this point can only be clarified in the light of the discussion of the related concept of collocation. For the moment, it is sufficient to point out that traditional accounts using syntactic and semantic criteria (and even not more radical accounts, such as Makkai’s) fail to include as ‘idiomatic’ such expressions as 2. “I have had it”.

4.2. MOTIVATION

Our motivation for revising the distinction between idiom and collocation lies in the recent development of corpus linguistics. By attempting to fix the applications of the terms phraseology, lexico-grammar and semantic prosody in relation to each other, we envisage a model of language in which phraseology embodies a continuum of expressions from pragmatically marked forms (idioms) to pragmatically unmarked or normal expressions (collocations). Unlike mainstream models of idiomatic expressions,
we use discourse criteria besides morpho-syntactic criteria to determine the idiomatic status of an expression.

This model presupposes that there are two forms of expression: a norm and a variant. This view of idioms and collocations had the advantage of being able to review the status of an expression (unmarked/marked) according to its context. The underlying assumptions are that a norm can be established and that the speaker has available to him or her, a variety of expressions, of which many can be identified as pragmatically marked. Any concept of core language (a concept of langue that assumes that peripheral forms are marked or special as in the term Languages for Special Purposes) appears to be incompatible with the fact that in any particular discourse, new norms are forged and become effectively the new core for that particular register or genre. We have seen that this new core can be particularly fixed and systematic, with unexpected symmetries, and this suggests that the underlying collocational patterns of a particular discourse are more prevalent and sensitive than they appear on the surface. The fact that certain discourses may be devoid of traditional idioms (as in science writing or Moon’s Bank of English) also suggests that rhetorical expressions are not absent, but that the new discourse has developed its own rhetorical devices.

In fact, Sinclair predicts that the so-called open expressions are more likely to occur in running text than canonical or stereotypical idioms (such as “it’s raining cats and dogs”).

4.3. CORPUS ANALYSIS

In order to improve upon the identification and transfer of idioms as shown in chapter 3, we need to carry out a computational analysis of text corpora in the light of the above observations. Similar analyses have enabled linguists earlier to search very large text archives systematically. There are two assumed advantages of computer-based corpus analysis: (i) as with the astronomer, the linguist can test theoretical armchair hypotheses by examining authentic data, and (ii) the size of the database can provide insights into language that had not been previously envisaged.
In particular, corpus analysis of idioms and the transformational or semantic properties of fixed expressions, rely towards the analysis of collocations and the distribution patterns of stereotypical phrases. For example, dictionaries now rely on corpus evidence, not only for the existence of words and phrases but, their use and distribution patterns (Mel chuk 1984, Sinclair 1991, Corrard and Grundy 1994).

In addition, corpus evidence on the distribution of idioms suggests that idioms are much less widespread and more variable than previously thought. Moon (1987), had found that of 2265 idioms (including a mixed category of metaphors) identified in the 323 million Word Bank of English (held at Birmingham University, England), 47% occur less than once per 4 million words. Of all the idioms examined, only 135 per million words occur more than twice (among these items Moon includes, “out of the blue”, “call the shots”, “foot the bill”). Moon concludes that pure idioms are somewhat marginal in nature, but are likely to be reformulated for stylistic effect (thus, “to be a penny short of a sixpence”, “to be a trunk short of a tree”). The original idiom becomes obscured, and all that is left is a framework: “to be an X short of a Y”.

4.4. METHODOLOGY

Native speakers usually have the natural ability to identify the idiomatic phrases from the regular sentences. However, to make the machine distinguish the idiomatic phrases from the regular sentences, we need to study the computational exploration of the text corpora. This study involves corpus analysis with reference to the idioms.

In this study, I would analyze each idiomatic phrase in comparison considering an additional word either preceding or following the idiom, till the sentence becomes self-explanatory that it is an idiomatic phrase.

Basing on this principle, I here examined the corpus that is available at the CALTS – Language Technology Lab, in the chapter of “Analysis of Idioms” and studied 3 million
words to propose certain rules for identifying idioms. However, the information required for identifying idioms is not always sufficient. Therefore, I take up in this chapter few idiomatic verbs for testing and evaluation on the basis of the ‘Argument Structure’ wherein the verb’s capacity or valency of arguments (thematic roles that the nouns carry with reference to the act or activity denoted by the verb is considered in order to prove their idiomaticity.

In order to study this, I have extracted data for ‘n-gram’ study (specific number of sequences of strings containing an idiomatic verb) from the corpus wherein an idiom is looked up within a contextual sequence of words. Here, the ‘n’ of the ‘n-gram’ stands for the number of strings (words) that either follow or precede an idiomatic phrase. I have increased this ‘n-gram’ string by including more number of strings either preceding or following the idiomatic phrase starting with one single word till I get to identify the particular phrase as an idiomatic phrase.

The expansion of ‘n-gram’ to ‘bi-gram’, ‘tri-gram’ and ‘tetra-gram’ was sufficient for me to identify all the idiomatic phrases that I have collected in my data. The ‘bi-grams’ have two word strings including one idiomatic verb. Similarly, the ‘tri-grams’ have three word strings including one idiomatic verb and the ‘tetra-grams’ have four word strings including one idiomatic verb and so on.

My study ends with the extraction of ‘tetra-grams’ strings incorporating extra three words that either preceding or following the idiomatic phrase expecting the required information in the identification of an idiom. The study of ‘n-grams’ is needed for this study for the identification of idiomatic phrases to gather information as a fool-proof method. In this case, some idioms are identified with a single word preceding the idiomatic verb. This word is usually observed to be either a noun or an adjective.
For example:

“vAdu gaddiwinnAdu”
“xiguluwo maMcaMpattu”
“sigguwo walavaMcu”, etc.

Here, “vAdu” is a pronoun and “xiguluwo”, “sigguwo” are adverbs. The additional information of noun in the case of ‘vAdu’, a [+human], is sufficient for identifying the idiomaticity of “gaddiwinu”. Similarly, for “maMcaMpattu”, the additional manner adverb “xiguluwo” preceding it is sufficient for identifying its idiomaticity. However, in the case of “walavaMcu”, it requires a manner adverb “sigguwo” preceding the idiomatic verb in order to identify its non-idiomaticity.

It has been observed that, there are many more such idioms that require similar and more information to identify their idiomaticity. For such cases, the study of ‘tri-grams’ is necessary for further evaluation of idioms.

For example:

“valalo padinAdani anukuMtArurA”

In this case, though the subject is [+hum], the idiom could not be identified due to the lack of the sufficient information to resolve the idiomaticity of the sentence. However, as the size of the ‘n-gram’ increases, as in the case of the ‘tetra-gram’, “mAyalAdi valalo padinAdani anukuMtArurA”, it resolves completely in identifying its idiomaticity. Here, the additional noun “mAyalAdi”, preceding the verb renders the complete idiomatic sense.

Again, in the case of “baruvu xiMcinatlanipiMciMxi”, a ‘bi-gram’ and “peVxxa baruvu xiMcinatlanipiMciMxi”, a ‘tri-gram’ could not give any additional information for the resolution of idiomatic sense. However, as the number of words in the string increases,
the possibility to get the additional information for the resolution of idiomatic sense is observed in the ‘tetra-gram’ as in “guMdeVllo peVxxa baruvu xiMcinatlanipiMciMxi”. Here, the extra information of “guMdeVllo”, a locative noun [-concrete], gives us the information required to resolve the idiomatic sense that was earlier lacking in the ‘tri-gram’.

Then, in the examples of “veleVwwi cUpu”, another ‘bi-gram’, and, “veleVwwi cUpataM welikakaxA”, a ‘tri-gram’, both lack sufficient information in the identification of an idiomatic phrase. However, “poVrapAtlu veleVwwi cUpataM welikakaxA”, a ‘tetra-gram’ gives us the complete information contextually due to the additional information of “poVrapAtlu”, a noun preceding the verb.

On the similar lines, “maMcaMpattu” [intr.v., “jabbuceyu”, ‘to become sick], a ‘bi-gram’ is not sufficient in the identification of an idiomatic phrase. Then the ‘tri-grams’, “maMcaMpatti maraNAvasWalo unna” or “awanu maMcaMpatti canipowU” have insufficient information and requires the ‘tetra-grams’, “maMcaMpatti maraNAvasWalo unna waMdriki” or “vqxXudEna awanu maMcaMpatti canipowU”. In this case, as the idiomatic verb requires the words like “maraNAvasWalo”, “vqxXudEna” etc., to identify the idiomaticity. Similarly, the following are some idiomatic verbs for testing and evaluating their idiomaticity:

4.5. EVALUATION

Here we would take up the study of some of the idiomatic fragments by running them against the available text corpus to observe the number of ‘n-grams’ required to resolve their idiomaticity. In the following examples, “@” is used to symbolize the sentences that help in the identification of an idiom.
1. Identifying the verb “bUjupattu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “pAwabadu”, to become outdated’, Hi. ‘saDa jAnA’

Consider some examples:

@AcArAlu bUjupattina sAMpraxAyAlu

@AcArAlu bUjupattina sAMpraxAyAlu vicCinna

@I kawwi bUjupattina BavAlaki

@aneka AcArAlu bUjupattina

anexi bUjupattina pAwa sixXAMwaM

@bUjupattina sAMpraxAlu rUpumAsi

@ bUjupattina BavAlaki punarjIvanam ivvatAniki

@ bUjupattina aBivqxxi niroXakamayina sAMpraxAyAnni

bUjupattina paxArWasevana valla xqRti

@ bUjupattina sAMpraxAyAlu vicCinnamayipovadaM

bUjupattina weVluguswrllu Barwalla iMtiperla

@iMglIRu praBAvaMwo walakAyalu bUjupattina
Identification:

(i) In these cases, it is observed that if the subject or Agent such as “AcArAlu”, “BavAlu” etc., or the complement of adjective that represent [-animate] or [-concrete], if either precedes or follows the idiomatic verb “bUjupattu”, then its idiomaticity is identified.

(ii) Similarly, in the example of “sAmpraxAyAlu bUjupattAy”, which could also be read as bUjupattina sAmpraxAyAlu”, wherein the “ina” form represents the adjectival form of “bUjupattu”, also renders an idiomatic sense if the subject or agent is [-animate] or [-concrete].

(iii) However, for identifying “bUjupattu” as a non-idiom, the subject should be locative noun of “ki” such as “puswakAlaki bUjupattiMxi” “iMtiki bUjupattiMxi”, or the adjectival forms such as “bUjupattina paxArWAAlu”, etc. wherein the subject always represents [-animate] or [-abstract].

(iv) In this case, although a ‘tri-gram’ may be sufficient to identify the idiomatic verb, a ‘tetra-gram’ always resolves the identification of the idiom.
2. Identifying the verb “baruvxiMcu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “bAXyawanirvarwiMcu”, ‘to fulfill responsibility’, Hi. ‘boJa uwAranA’

Consider some examples:

baMdeVku baruvu xiMcinatlayi gattigA

@guMdeVlo peVxxa baruvu xiMcinatlanipiMciMxi

baruvu xiMcinatlayi gattigA nittUrpu

**Identification:**

(i) From the above examples, the verb “baruvuxiMcu” occurs with the locative form of “lo” as in “guMdeVllo baruvu xiMcinatlanipiMciMxi”, it would always render an idiomatic sense.

(ii) However, if the subject or agent is [-animate] or [-abstract], as in the case of “I saMciloni baruvu xiMcu”, “nA wala pEna unna I mUta wIsi baruvu xiMcu” wherein the locative forms such as “loni”, “pEna”, occur with the complement nouns [-abstract], then these would always be in a non-idiomatic sense.

(iii) From the above examples, only a ‘tetra-gram’ that usually resolves the idiomatic verb and helps in the identification of the idiom.

3. Identifying the verb “ceyicAcu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “AsiMcu”, ‘to aspire’, Hi. ‘azcala pasAranA’
Consider some examples:

catukkuna ceyi cAci kAxammA

@mIrU oVsaganunna rAjamakutamunaku ceyicAcuvAru

@oVsaganunna rAjamakutamunaku ceyicAcuvAru eVvarunuleru

wrosesinatlu catukunna ceyi cAciMxi

Identification:

(i) It is observed that if the subject or agent has a dative marking “ku” as part of its argument structure and that always precedes the idiomatic verb, then it would always render an idiomatic sense.

(ii) In this case, a ‘tetra-gram’ that usually resolves the idiomatic verb and helps in the identification of the idiom.

4. Identifying the verb “gUdukattukoVnu” in an idiomatic usage.

Nature: intransitive verb
Meaning: “nilaciMdu”, ‘to stay put’, Hi. ‘GoMsalA banAnA’

Consider some examples:

A medalo gUdu kattukoVnAniki

@Awma keMxraKawvaM gUdu kattukoVni

gUdu kattukoVnAniki oVka piccikEnArAxu
Identification:

(i) If the idiomatic verb “gUdukattukoVnu” have [+hum] or [+ani] as its subject, and occurs either preceding or following the idiomatic verb, then it will always trigger a non-idiomatic sense.

(ii) If the subject is [-animate], then it would always be an idiomatic verb.

(iii) If the subject is [-concrete] with the indicator of locative clause “lo” as in the case of “guMdeVllo”, etc., then it would always trigger an idiomatic sense. “guMdeVlu (pl.)” is identified as a different noun meaning ‘mind’ as against the word “guMdeV”, eg. ‘heart’.

(iv) Here a ‘tetra-gram’, helps in the identification of the idiom.

5. Identifying the verb “gaddiwinu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “awyAsapadu”, ‘to become greedy’, Hi. ‘Gusa KA’

Consider some examples:
@bAkAruku veVduwU buxXi gaddiwinini

@buxXi gaddiwinini I mate

@buxXi gaddiwinini A lAtarI

@buxXi gaddiwinini pekAta xagXara

@buxXi gaddiwinnaxi yixigo leVMPalulu

@buxXi gaddiwinini palakariMcAnani

dabbu kosaM gaddi winAlA

gaddi winAlA eMti

gaddi winAdAniki sixXaMgA uMxi

gaddiwinini A lAtarI tikkeVtlu

gaddiwinini pekAta xagXara jeri

meka gaddi winAdAniki sixXaMgA

@nA buxXi gaddiwinini

nAnA gaddiwinini wIsukoVcciveswe mUrIAlA wagaleswAremite

nenu nAnA gaddiwinini wIsukoVcciveswe mUrIAlA

@peji 25 buxXi gaddi wini
Identification:

(i) In this case, if the subject for the verb “gaddiwinu” is either [+abstract], then it would always trigger an idiomatic verb.

(ii) If the subject is [+animate] and [-human], such as “meka”, “Avu”, etc., then it always is a non-idiomatic verb.

(iii) In this case, a ‘tetra-gram’ helps the idiomatic verb in the identification of an idiom.

6. Identifying the verb “goVMwukoyu” in an idiomatic usage:

Nature: intransitive verb.
Meaning: “anyAyaMceyu”, ‘to do injustice’, Hi. ‘galA kAtanA’

Consider some examples:

Ame goVMwu kosi caMpeVnu

ani AmeV goVmwu kosi

anucu wanagoVmwu koyamani Barwanu

goVMwu kosukoVni nIlIYIYuleni bAvilo

goVMwukosinAdu xositlo rakwamu pattinAdu

gogunArawo goVMwu kosukoVni nIlIYIYuleni
karakara karawakawwiwo goVMwukosinAdu xositlo

mosagAdu ImeV goVMwu koyalexu

@o vyasanaparudikicci xInigoVMwu kosAru AvexanagA

wama goVMwu koyamani Barwanu koreVnu

xInigoVMwu kosAru AvexanagA ceVppiMxi

xuswulavAlYIYu wana goVMwu koswunnAru

**Identification:**

(i) In this case, if the verb “goVMwukoyu”, in the context of a subject occurs with the instrumental “icci”, then it would always trigger an idiomatic verb.

(ii) In this case, a ‘hexa-gram’ helps in the identification of the idiomatic usage.

7. Identifying the verb “goVdugupattu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “maxxawu cUpu”, ‘to assist’, Hi. ‘CawarI lagAnA’

Consider some examples:

A goVdugu patta valeVnu

pillavAdiki padageVwwi goVdugu pattiMxata

dillImIxa varRaMpadakuMdA goVdugu pattukuMxannAdu
goru vaMkaku goVdugu pattexi

rAGava BowikakAyAniki goVdugu pattAli

vAdiki manaM goVdugu pattAli

vaMkaku goVdugu pattexi eVvaru

varRam vaswuMxemo goVdugu pattukeVIYYaMdi

varRaM padakuMdA goVdugu pattukuMxannAdu

I yeVMdaku A goVdugu pattAli

**Identification:**

(i) As seen above, though there are no related examples found from the corpus, it can be concluded that if the subject for the verb “goVdugupattu” is preceded by the special usage of “e eVMda”, followed with the dative preposition “ku” or “ki”, then it would always trigger an idiomatic sense.

(ii) Here, a ‘hexa-gram’ helps in the identification of the idiomatic usage.

8. Identifying the verb “gowuluwavvu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “Apaxalu weVccu” ‘to dig pitfalls’, Hi. ‘gadDe KoxanA’

Consider some examples:

gowulu wavve Past assisteVMtu
Identification:

(i) If the verb “kAlujAru” is suffixed by the lexical terms such as “kriMxa” or “venaka”, then it would always be an idiomatic usage.

(ii) If the subject or agent has a dative marking “ki” as part of its argument structure and that always precedes the idiomatic verb, then it would always render an idiomatic sense.

(iii) In this case, a ‘tetra-gram’ helps in the identification of the idiom.


Nature: intransitive verb.

Meaning: “sllaMpovu”, ‘to lose virginity’, Hi. ‘pAzva PisalanA’

Consider some examples:

kAlujari neVlanupadi

A prayawnaMlo aprayawnaMgA kAlujAri
RikAruki veVIYlYikAlu jAri

aMxuvalla kAlujAri caMxramma bAvilo

aprayawnaMgA kAlujAri bAvilo paddAdu

buraxalo kAlu jArinatlugA ayi

cIma kAlu jAri pAyasaMlo

kAlu jAri nItlo padi

kAlu jari nelanupadi

kAlu jAri pAyasaMlo padipoyiMxi

eVkkuwU uMdagA kAlujAri nAlugExu

kAlujAri caMxramma bAvilo padipoyiMxi

kAlujAri nAlugExu meVtlamIxa nuMdi

kAlujAri vaMweVna mIxanuMdi kAluvalo

kRaNaMlo AmeV kAlu jAriMxe

kalipina yuvakudu woVlisArigA kAlujArina

@mAhasika xOrbalyaM valla kAlujArina
meda eVkkuwU uMdagA kAlujAri

@*OrbalyaM valla* kAlujArina maguva

xigucuMdagA humAyUn kAlujAripadi maraNiMceVnu

**Identification:**

(i) In this case, if the verb “kAlujAru” is either preceded or followed by the locative nouns such as “nIlo”, “buraxalo” “pAyasaMlo”, etc., then it would always be a non-idiomatic usage.

(ii) If the verb “kAlujAru” is either preceded or followed by the noun “xOrbalyaM”, then it would always be an idiomatic usage.

(iii) In this case, a ‘tetra-gram’ helps in the identification of an idiom.

10. Identifying the verb “**kannumUyu**” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “canipovu”, ‘to die’, Hi. ‘xama wodanA’

Consider some examples:

APIsaru kudi kannu mUsukuni

axi guddi kannu mUsinA

ceVppinatlu kannumUyadaM atuvaMtixi kAxu

prasaviMci kannu mUsiMxi

gAMXlIjI menakodali Barwa kannumUsAdu
gaddamIxa kannu mUyAlanna korkeV

guddi kannu mUsinA oVkate wericinA

@janavari 1943 na kannumUsAru

kAni anna kannu mUswe

kAni pEna ceVppinatlu kannumUyadaM

kannu mUsi eVMwakAlamEMxo

kannu mUsina xAcukoVnnavAru eVccata

kannu mUswe ika AxaraNa

kannu mUwapadina A rAwri

kannu mUya emiyU kanabadaxu

kannumUsi oVkka pAvugaMtEnA gAlexu

kannu mUsinA oVkate weVricinA oVkate

kannu mUswe cAlu A pAmu

kapila maharRi kannu mUsi

koVxxi kRaNAlu kannu mUswAdu
mI kannu mUyumú

maharRí kannu mUsí eVMwakAlamEMxo

mana jFAnexMxriyamulu kannu mUyu

marunAtí pagalu kannu mUwapadáni

nA kumArunikivvu ani kannumUsAdú

oVkaru kannu mUsí koVMxaru

pEna cEvpinatlu kannumUyadaM atuvaMtiixi

pagalu kannu mUwapadáni

porí raNaraMgamuna kannu mUsiri

@punnammagAru 1920 lo kannumUsiMxi

vaccina guMdeVpotuwo kannu mUsAdú

weVlugu gaddamIxa kannu mUyAlanna

xInilo oVkaru kannu mUsí

**Identification:**
Here, if the verb “kannumUyu” is preceded by the post-position “lo”, “na”, etc., that indicates [+adverb] spatial and temporal, then it would always trigger an idiom.

In this case, a ‘tetra-gram’ helps in the identification of an idiom.

11. Identifying the verb “maMcaMeVkku” in an idiomatic usage.

**Nature**: intransitive verb.

**Meaning**: “jabbupadu”, ‘to become sick’, Hi. ‘cArapAI pakadanA’

Consider some examples:

appudu parIkRallo mIru maMcaMeVkki

kAlYlu kadukkuni maMcaM eVkkAdu

kadukkuni maMcaM eVkkAdu accayya

laMboxaraM maMcaM eVkkAdu kAnI

mIru maMcaMeVkki parIkRanu ceVtteVkkiswAru

maMcaM eVkkAdu kAnI gummaM

maMcaM eVkka boye muMxu

maMcaM eVkke muMxugAne l

@nenu rogaMwo maMcaM eVkke

padukoMdi laMboxaraM maMcaM eVkkAdu
parlkRallo mIru maMcaMeVkki parlkRanu

rAjU maMcaM eVkka boye

rôgaMwo maMcaM eVkke muMxugAne

**Identification:**

(i) In this case, the verb “maMcaMeVkku” in the context of lexical terms such as “rôgaM”, “jabbu”, “xigulu”, “beVMga”, and such similar words indicating the psycho-semantic states, when occurs with the instrumental “wo”, then it would always trigger an idiom.

(ii) In this case, a ‘tetra-gram’ helps in the identification of an idiom.

12. Identifying the verb “maMcaMpattu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “jabbupadu”, ‘to become sick’, Hi. ‘cArApAI para padanA’

Consider some examples:

xiguluwo maMcaM pattakaMdi

awanu maMcaMpatti canipowU nuluguru

@boVggu pariSrana haTAwwugA maMcaMpattiMxi

cewwo walamIxa nulaka maMcaM pattukunnaAdu

exI yI maMcaM pattaMdi
Identifying the verb “nadumubigiMcu” in an idiomatic usage.

Nature: intransitive verb.

Meaning: “moxalupeVttu”, ‘to begin’, Hi. ‘kamara kasanA’

Consider some examples:

@mawaM nadumu bigiMci raMgaMlo

nadumu bigiMci kqRi sAgiswunna

nadumu bigiMci raMgaMlo xiginaxi
Identifying the verb “raMgubayatapadu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “pravarwana weliyu”, ‘to become exposed’, Hi. ‘kalayI KulanA’

Consider some examples:

@wodelu asalu raMgubayatapaduMxi

**Identification:**

(i) Here, if the verb “raMgubayatapadu”, is preceded by the noun that is [+animate] [-concrete], then it would always trigger an idiom.

(ii) In this case, a ‘tetra-gram’ helps in the identification of an idiom.

15. Identifying the verb “valalopadu” in an idiomatic usage.
**Nature:** intransitive verb.

**Meaning:** “maMwramugXudu avvu”, ‘to be charmed’, Hi. ‘caMgula meM PazsanA’

Consider some examples:

@eVvaro mAyalAdi valalo padinAdani

@mAyAlAdi valalo padinAdani anukuMtArurA

manaM xAni valalo padipowAM

ValYlYu valalo padataM I

valalo padataM I mosware

valalo padinAdani anukuMtArurA kAbolu

wakkuva vAlYlYu valalo padataM

**Identification:**

(i) If the verb “valalopadu” is preceded by the noun “mAyalAdi” that is [+human] [-concrete], then it would always trigger an idiom.

(ii) In this case, a ‘tetra-gram’ helps in the identification of an idiom.

16. Identifying the verb “veleVwwicUpu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “niMxiMcu”, ‘to blame’, Hi. ‘uzgall uTAnA’
Consider some examples:

batti vixyArxulaku veleVwwi cUpi

@cesina poVrapAtlu veleVwwi cUpataM

@eVtuvEpu wappuMte atuvEpu veleVwwi cUpadame kavilakRyaM

@poVrapAtlu veleVwwi cUpataM welikakaxu

veleVwwi cUpi lABaM lexu

vixyArxulaku veleVwwi cUpi lABaM

**Identification:**

(i) In this case, if the subject or Agent such as “poVrapAtu”, “wappu”, etc., that represent [-animate] or [-concrete], either precedes or follows the idiomatic verb “veleVwwicUpu”, then it would always trigger an idiom.

(ii) In this case, a ‘tetra-gram’ is sufficient in the identification of an idiom.

17. Identifying the verb “wAtAkulukattu” in an idiomatic usage.

**Nature:** intransitive verb.

**Meaning:** “amAyakudinicyu”, ‘to fool’, Hi. ‘ullU banAnA’

Consider some examples:

@ceyiMcukuni vArike wAtAkulu kattattaM

muvva ani wAtAkulu kattAru
Identification:

(i) If the subject occurs with the dative noun “ki” that is [+hum] [-concrete], that either precedes or follows the verb “wAtAkulukattu”, then it would always trigger an idiom.

(ii) Here, a ‘tri-gram’ is sufficient in the identification of an idiom.

18. Identifying the verb “walavaMcu” in an idiomatic usage.

Nature: intransitive verb.
Meaning: “loVMgu”, ‘to submit’, ‘subdued’, Hi. ‘sara JukAnA’

Consider some examples:

@rAju vinayaMwo walavaMci

@nAku walavaMcani vlrudu

@A slwammakowi vicAraMgA walavaMcukoVni

@rAjyAXikArAnni prajalu gurwiMci xAniki walavaMcAru

@AgrahAveSAlaki kUdA oVkkoVkkappudu walavaMcAli

AmeV walavaMci gumisi wappu
AmeV vaMka cUsi wala vaMcesAdu

SqMgAramunu varNiMcetappudu AmeV walavaMci

@aMxuku sigguwo walavaMcukoVni veVMtane

anukUdani mAta annAnemonani wala vaMcuukunnAnu

anumAnaM vacce waruvAwa walavaMci

awadu wala vaMcuuki neVmmaxigA

awani veVMbade walavaMcukoVni ceVnnudu

@buxXudEnatlu walaci viXiki walavaMci

@brahmanAyudu avamAnamuce walavaMcukoVni niluabadeVnu

naMganAcilA walavaMcukoVni golYlYugillu

emI eVragani naMganAcilA walavaMcukoVni

ippudu wala vaMcuwunnAnu annAru

@ixi mAnavawvaM walavaMcukovAlsina viRayam

kUdA oVkariki wala vaMcunu

kUrcuni bidiyaMgA wala vaMcuukuMxi

@kaRtanaRtAlaku wala vaMcaka kqRi
ketAyimcina paxavini walavamcukuni

lajjagA wala vaMcukuMxi

leci meVlagA wala vaMcukuni

@mA vAsaSWanamu sigguce walavaMci

mIsAnni peVrukuMtU awadu walavaMci

malYIYI walavaMci aparAXaM cesinavAdilAgA

@nAku walavamcani vIrudu eVVvaru

nAluggaMtalaki awanu wala vaMcukoVni

nenu aMxarAgAni walavaMcuoVni nadavAlsiMxe

@pAwarakaA jIviwAlake walavaMci vuMtAru

@pOrudu cattAlaku walavaMcuwAdu

@rAjlPadi walavamcuku jIviMce sWiwini

@rAjakIya sArvaBOmAXikArumunaku walavamcavaleV

sigguce walavaMci xuHKamuce kumilipoyiMxi

siggwo walavamcukoVni brawuku
Identification:

(i) In this case, if the subject occurs with the dative noun “ki” that is [-human] [-concrete], that precedes the verb “walavaMcu”, then it would always trigger an idiom.

(ii) If the verb “walavaMcu”, is either preceded or followed by the manner adverbs such as “sigguwo”, “vicAraMwo”, “vicAraMgA”, “bidiyaMwo”, “avamAnamuce”, “naMganAcilA”, “lajjawo”, and such similar words indicating the psycho-semantic states, “ce, “wo”, then they would always trigger non-idiomatic sense.

(iii) If the verb “walavaMcu”, is preceded by nouns functioning as subjects [+abstract] such as “mAnavawaM”, “hewuvAxaM”, “palukubadi”, etc., then it would always trigger an idiom.

(iv) Here, a ‘tri-gram’ is sufficient in the identification of an idiom.

19. Identifying the verb “walawirugu” in an idiomatic usage.

Nature: intransitive verb.

Meaning: “viciwriMgA vuMdu”, ‘to be eccentric’, ‘head reeling’, Hi. ‘hoSa udajAnA’

Consider some examples:

walanoVppi wala wirugudu uMtAyi

@walawirugudu  panulu cesAvo iMtiki

@walawirugudu  panulu ceswAdani anukuMtunnAdu

walawiruguta vaMti xoRamulu nivAriMca
Identification:

(i) In the example of “walawirugudu panulu”, wherein the “du” form represents the nominalization of “walawirugu”, renders an idiomatic sense if the subject or agent is [-animate] or [-concrete].

(ii) Here, a ‘tetra-gram’ helps in the identification of an idiom.

4.6. CONCLUSION

The study is restricted to only few idioms for exemplification. It is an effort to explore the possibilities of building a tool, pragmatic identification and transfer of sense in Machine Translation. In the process we have explored the conventional morpho-syntactic information for resolving the idiomatic sense from non-idiomatic sense. The current chapter explored the possibilities of priority at more reliable information for the resolution of idioms. Therefore, to lay our hands on the additional information, we have used ‘n-gram’ technique. The basic idea underlying this method is to find an answer to a question, how large a string/sequence of words should be optimum to be resolved as an idiom.

The observations from the ‘n-gram’ study of the selected idioms to resolve their idiomatic sense show that the maximum confidence to resolve a sequence as an idiomatic structure requires: (i) at least a set of four words, (ii) the idiomatic verbs are usually preceded by either nouns or adverbs as clue words, (iii) if the idiom is an adjectival form of verb, then the minimum length the sentence structure requires for identification of that particular idiomatic verb requires either one word preceding or following an idiom for that particular sentence, (iv) if the idiom is in a verbal form, then the minimum length required for that particular sentence should at least have three words preceding that particular idiomatic verb.
For optimum resolution, unlike mainstream model of idiomatic expressions, therefore, we can use discourse criteria to determine the idiomatic status of an expression though many idiomatic expressions still require further analysis and identification criteria to resolve their idiomaticity.