Acta Linguistica Asiatica

Volume 9, Issue 1, 2019

ACTA LINGUISTICA ASIATICA
Volume 9, Issue 1, 2019

Published by: Znanstvena založba Filozofske fakultete Univerze v Ljubljani
(Ljubljana University Press, Faculty of Arts)

Issued by: Department of Asian Studies

For the publisher: Dr. Roman Kuhar, Dean of the Faculty of Arts

Journal's web page:
http://revije.ff.uni-lj.si/ala/
The journal is published in the scope of Open Journal Systems

ISSN: 2232-3317

Abstracting and Indexing Services:
Scopus, COBISS, dLib, Directory of Open Access Journals, MLA International Bibliography, Open J-Gate, Google Scholar and ERIH PLUS.

Publication is free of charge.

Address:
University of Ljubljana, Faculty of Arts
Department of Asian Studies
Aškerčeva 2, SI-1000 Ljubljana, Slovenia

E-mail: nina.golob@ff.uni-lj.si

TABLE OF CONTENTS

FOREWORD

In the mids of cold northern winds and landscape covered with snow we are pleased to announce the first ALA issue of the year 2019, which contains six research articles. Warm congratulation goes to all the authors, and words of appreciation to the Editorial team and recently enlarged proofreading team that have been working very hard in order to offer state-of-the-art contemporary linguistic research in this journal.

The present issue is opened up by **Mayuri J. DILIP** and **Rajesh KUMAR**, who present a unified account of licensing conditions of Negative Polarity Items (NPI) in Telugu. In their work "Negative Polarity Items in Telugu" they analyze the distribution of NPIs in complex sentences with embedded clauses, and conclude that negation c-commanding NPI be conducted at the base-generated position.

Kun SUN with his article "The Integration Functions of Topic Chains in Chinese Discourse" thoroughly presents the long and extensive Chinese research tradition on topic chains, and re-examines their core characteristics with the help of the so-called "integration functions".

The following paper "Tracing the Identity and Ascertaining the Nature of Brahmi-derived Devanagari Script" by **Krishna Kumar PANDEY** and **Smita JHA** exploits the orthographic design of Brahmi-derived scripts. Authors argue that such scripts should not be described with the existing linguistic properties of alphabetic and syllabic scripts but should instead gains its own categorization with a unique descriptor.

Chikako SHIGEMORI BUČAR successfully submitted the article "Image of Japan among Slovenes" in which she represents the process and mechanism of borrowing from Japanese into Slovene. Conclusions briefly touch the image of Japan seen through the borrowing process and consolidated loanwords, and predict possible development of borrowing in the near future.

Another interesting paper "Understanding Sarcastic Metaphorical Expression in Hindi through Conceptual Integration Theory" was authored by **Sandeep Kumar SHARMA** and **Sweta SINHA**. Based on a corpus of five thousand sentences, authors examine the abstract notion of sarcasm within the framework of conceptual integration theory, and with special reference to Hindi language. Findings aim to provide a theoretical understanding on how Hindi sarcasm is perceived among the native speakers.

And last but not least, **Điệp Thi Nhu NGUYỄN**, **An-Vinh LƯƠNG**, and **Điền ĐINH** humbly observe research backlog in the area of Vietnamese text readability and write their paper "Affection of the part of speech elements in Vietnamese text readability" to

encourage researchers to further explore the field and put Vietnamese findings on the world's map.

Editors and Editorial Board wish the regular and new readers of the ALA journal a pleasant read full of inspiration.

Editors

RESEARCH ARTICLES

NEGATIVE POLARITY ITEMS IN TELUGU

Mayuri J. DILIP
Indian Institute of Technology Madras, India
mayuri.dilip@gmail.com

Rajesh KUMAR
Indian Institute of Technology Madras, India
thisisrajkumar@gmail.com

Abstract

The paper presents a unified account of licensing conditions of Negative Polarity Items (NPI) in Telugu. Based on the distribution of NPIs in complex sentences that consist of embedded clauses, we state that negation c-commanding NPI is at the base-generated position. Consequently, features checking between negation and NPI restricts the alternatives on the scale inherent to NPIs. The morphological realization of NPI in the non-negative contexts is different from the context with overt negation. The NPIs show the following distribution. NPI occurs in subject position; A negative licensing Multiple NPIs. There are three types of NPIs: wh-element, quantifier and idiomatic expression. In complex sentences, wh-elements block long-distance licensing. In contrast, quantifiers and idiomatic expressions do not block long-distance licensing.

Keywords: Negative Polarity Item; minimalist-based approach; feature checking; quantifier scale; c-commanding

Povzetek

Članek predstavlja celovit pregled pridobitvenih pogojev (angl. licencing conditions), ki zadevajo k nikalnosti usmerjene izraze (Negative Polarity Items ali NPI) v teluščini. Na osnovi porazdelitve teh izrazov v sestavljenih stavkih z vrinjenimi stavki zagovarjamo tezo, da se s-poveljevanje k nikalnosti usmerjenim izrazom izvede na položajih, ki izhajajo iz osnove. Posledično preverjanje značilnosti med nikalnostjo in k njej usmerjenimi izrazi omejuje druge možnosti in sicer preko lestvice, ki je povezana z izrazi NPI. Morfološka realizacija takšnih izrazov v nenegativnih kontekstih je drugačna od kontekstov z očitno negacijo. K nikalnosti usmerjeni izrazi izkazujejo naslednjo porazdelitev. Pojavljajo se na položaju osebka kot negativno pogojeni večkratni izrazi NPI. Obstajajo trije tipi k nikalnosti usmerjenih izrazov: vprašalnice, števniki in

Acta Linguistica Asiatica, 9(1), 2019.
ISSN: 2232-3317, http://revije.ff.uni-lj.si/ala/
DOI: 10.4312/ala.9.1.9-28

idiomatični izrazi. V sestavljenih stavkih vprašalnice zaustavijo oddaljeno pridobivanje, medtem ko ga števniki in idiomatski izrazi ne.

Ključne besede: k nikalnosti usmerjeni izrazi; minimalistični pristop; preverjanje značilnosti; lestvica števnikov; s-poveljevanje

1 Introduction

This paper discusses the syntactic description of Negative Polarity Items (NPI) in Telugu, a Dravidian language. An NPI usually requires a negative licensor as discussed in several studies such as Lasnik (1972), Linebarger (1980), and Laka (1989), Progovac (1994) among others. Some argue for overt licensing involving NPIs (Lasnik, 1972; Kumar, 2006) whereas others argue for licensing at some other level such as Logical Form (LF) (Line Barger, 1980; Laka, 1989; Mahajan, 1990; Progovac, 1994; Balusu et al., 2016). This paper aims at providing a unified account of licensing conditions of NPIs in Telugu, specifically wh-elements (wh-NPI), quantifiers (q-NPI) and idiomatic expressions (i-NPI), in negative & non-negative contexts and in local & long-distance licensing contexts by adopting Kumar's (2006) analysis which is further modified into a minimalist-based approach. In order to depict feature checking between negation and NPI, we adopt operation Agree (Chomsky, 2000) and scalar reasoning (Chierchia, 2013; Nicolae, 2012).

The organization of the paper is as follows: section 2 discusses the basic structure of negation and affirmation; section 3 discusses the structural distribution of wh-NPIs, q-NPIs and i-NPIs; NPI in subject position; the occurrence of multiple NPIs, the complement clauses exhibiting restriction on long-distance licensing of NPIs; NPIs in non-negative contexts; section 4 illustrates the quantificational structure of NPIs; section 5 describes licensing conditions of NPIs; section 6 is conclusion.

2 Structural description of negation and affirmation

The morphological and syntactic description of negation and affirmation is necessary, since they play a significant role in restricting the distribution of NPIs in Telugu. The discussion is elaborated below.

In Telugu, the morphological structure of sentential negation varies in verbal predicate with a content verb as in (1a) and a non-verbal predicate such as an existential verb as in (1b). In every construction including (1a) and (1b), the negation always precedes the agreement marker. Where the agreement marker also functions as a finiteness marker in Telugu. In the case of the verbal predicate as in (1a), the negative marker *lē* occurs as a bound morpheme suffixed to the main verb except in the case of the verb with future tense. In future tense, an overt form of negative marker is absent. Therefore, we assume that the negative marker occurs in the same position

as in past and present tense and we indicate its presence with ø 'a zero morpheme'. There are various negative markers such as *-vaddu, -kūḍadu, -a-, -aka-, -akunḍā/-kunḍā, -aku-, lē-* and *ani-* (Krishnamurti & Gwynn, 1985). These markers occur in different contexts and they follow the main verb similar to the marker *–le*. In the case of non-verbal predicate as in (1b), the negative occurs as a fusional morpheme, when the verb is in the present or the past tense. By fusional morpheme, we mean that the verb functions both as negation as well as copula. However, in future tense, it occurs as a bound morpheme suffixed to copula *un-* 'be'. In (1a) and (1b) the past tense and the present tense morphemes are homophonous.[1]

1a. rāmu rā-**lē**-du/ rā-**lē**-du/ rā-ø-ḍu
Ramu come-PRES.**NEG-**3.SG.N/come-PST.**NEG-**3.SG.N/come-FUT-**NEG-**3.SG.N
'Ramu does not come. / did not come. / will not come.'

1b. rāmu inṭi-lo **lē**-ḍu/ **lē**-ḍu/ unḍ-**a-**ḍu
Ramu house-LOC be.PRES.**NEG-**3.SG.M/be.PST.**NEG-**3.SG.M/be-FUT.**NEG-**3.SG.M
'Ramu is not at home./ was not at home./ will not be at home.'

Lahiri (1998), Kumar (2006) and Bhattacharya (2012) show that the element that occurs as an NPI also occurs in certain non-negative contexts. In section (3.5), (non-)negative contexts show that the morphological structure corresponding negative contexts never occur with a non-negative contexts. In contrast, the morphological structure corresponding non-negative contexts never occurs in a negative context. The point to be noted is that the variations correspond to the functional categories such as negation and affirmation. Therefore, the formal negation and affirmation play a role in determining the type of NPI. Before moving on to how the functional categories license NPIs, we describe the syntactic representation of negation and affirmation in a tree structure. In a syntactic representation, similar to the previous studies such as Pollock (1989)**,** Chomsky (1991), Mahajan (1990), Kumar (2006), Laka (2016), the negative heads

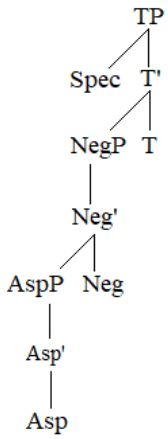
[1] The tense being past or present can be proved by positing the time adverb *ninna* 'yesterday' or *ippuḍu* 'now' in (1a) and (1b). In the presence of *ninna* 'yesterday', the verb exhibits past tense. In contrast, in *ippuḍu* 'now' occurs in (1b), then the verb indicates present tense.

i a. rāmu {ninna/ippuḍu} rā-**lē**-du
Ramu {yesterday/now} come-{PST/PRES}**.NEG-**3.SG.N
'Ramu did not come yesterday.'/'Ramu will not come now.'

ii b. rāmu {ninna/ippuḍu} inṭi-lo **lē**-ḍu
Ramu {yesterday/now} house-LOC be-{PST/PRES}.**NEG-**3.SG.M
'Ramu was not at home yesterday.'/'Ramu is not at home now.'

its own phrase called NegP and the NegP projects below TP and above AspP, located in the functional domain as in (2).

2. Position of negation in a tree structure:



In the case of affirmation in the non-negative contexts, we adopt Laka's (2016) analysis, where the affirmation heads its own syntactic projection called an Affirmative Phrase (AffP) and it occurs in place of NegP. Morphologically, there is no overt marker for affirmation in Telugu, hence we posit a zero morpheme exactly in the same position where a negative morpheme occurs in a negative clause. Now, in the following section, we discuss the structural description of NPIs in Telugu by comparing it with English and Hindi.

3 Structural description of NPIs

In this section, we demonstrate the variations and restrictions of NPIs occurring in different types of clauses. The structural variations that we discuss in this section help us understand the licensing conditions of NPIs. We observe the structure of Telugu by comparing it with Hindi and English. Such comparison among languages belonging to three different language families helps us to identify the location of the licensing conditions in the clause structure.

3.1 Description of three types of NPIs

In line with Lahiri (1998), Kumar (2006) and Balusu et al. (2016) the NPIs are attached with a particle indicating 'even'. In Telugu, the particle is *kūḍā* or the final vowel lengthening. The root of an NPI can be a wh-element, quantifier or an idiomatic expression similar to Hindi as in (3), (4) and (5). To the best of our knowledge, we do not find a wh-element and a quantifier as NPIs in English. However, English has idiomatic expression as in (5c).

3a. *Telugu*
rāmu **evvarini-i** {avamāmāninc-a-lē-du/*avamāmāninc-ā-ḍu}
Ramu **wh-even.NPI** {insult-PST-NEG-3.SG.N/insult-PST-3.SG.N}
'Ramu did not insult anyone.'

3b. *Hindi*[2]
rāmu ne **kisī kā bhī** apmān {nahĩĩ kiyā thā/ *kiyā thā}
Ramu erg **wh even-NPI** insult {not do PST.SG/ do PST.SG}
'Ramu did not insult anyone.'

4a. *Telugu*
rāmu **okkarini kūḍā** avamāninca-lē-du/avamāninc-ā-ḍu
Ramu **one person even.NPI** insult-NEG.PST-3.SG/ insult-PST-3.SG
'Ramu did not insult anyone.'
Literally: 'Ramu did not insult one person also'

4b. *Hindi-Urdu*
rāmu ne **ēk vyakti ka bhī** apmān nahi kiyā thā/ *kiya thā
Ramu erg **one person gen even.NPI** insult not do PST.SG/ do PST.SG
'Ramu did not insult anyone.'
Literally: 'Ramu did not insult one person also'

5a. *Telugu*
nēnu nī-ku **cilli gavva kūḍā** {ivv-a-nu/*ist-ā-nu}
I you-to **single penny even.NPI** {give-NEG.PST-3.SG/*give-PST-3.SG}
'I will not give a single penny to you.'

5b. *Hindi*[3]
mẽ tum-kō **ēk phūṭi kauṛī** {nahĩĩ dū-ngā/ *dū-ngā}
I you-dat **one broken penny.NPI** {NEG give-FUT.1.SG.M/ give-FUT.1.SG.M}
'I will not give you a red cent.'

5c. *English*
He did not save **a single penny**.
(Ladusaw, 1983, p. 382; as cited in Ramachandram, 1991, p. 13)

3.2 Occurrence of NPI in subject position

Below are the constructions exhibiting the NPI in subject position. Both in Telugu and Hindi, all the three types of NPIs can occur in subject position as in (6a) and (6b). In English, the NPI *anyone* cannot occur in subject position as in (6c), since it cannot be licensed by structurally lower negation (Chierchia, 2013, p. 62).

[2] Personal communication with Dr. Devleena Chakravarty, Ph.D.

[3] Kumar (2006).

6a. *Telugu*
ēmī/okkaṭi kūḍā/cilli gavva kūḍā lē-du
wh-NPI/q-NPI/i-NPI be.PRES.NEG-3.SG.NH
'There is nothing.'

6b. *Hindi*
kōi bhī/ ēk bhī/ ēk phūṭi kauṛī nahĩĩ hẽ
wh-NPI/q-NPI/i-NPI not be.PRES.SG
'There is nobody.'/ 'Not even one person is there.'/
'Even a single penny is not there.'

6c. *English*
***Any** student didn't respond well. (Chierchia, 2013, p. 62)

3.3 Occurrence of Multiple NPIs

Below are the constructions with multiple NPIs licensed by a single negative licensor. In English as in (7a), the negative *not* licenses the NPI such as *anybody, anything*; in Hindi as in (7b), the negative *nahĩĩ* licenses the NPIs *kōī bhī* 'anybody', *kahī* 'anywhere' and in Telugu as in (7c), the negative *-lē* licenses the NPIs *evarū* 'anyone'*, ekkadikī* 'anywhere'. Note that i-NPI cannot occur as multiple NPIs.

7a. *English*
He didn't give **anybody anything** at **any** place at **any** time.
(Kuno & Whitman, 2004, p. 225)

7b. *Hindi*
kōī bhī **kahī** nahĩĩ Gayā
who even.NPI **somewhere.NPI** not go-PST.3.SG.M
'Nobody went anywhere.'

7c. *Telugu*
evarū **ekkadikī** vell-a-lē-du
who.even.NPI **anywhere.NPI** go-PST-NEG-3.SG.N
'Nobody went anywhere.'

3.4 Structural restrictions of NPIs in various complex sentences

In this section, we discuss the distribution of NPIs occurring in three types of complex sentences: adjunct clauses, complex NPs and complement clauses. Based on the structural restrictions in complex sentences, we classify the q-NPIs and i-NPIs as non-strict NPIs and wh-NPIs as strict NPIs. A strict NPI is the one which does not permit long-distance licensing and in contrast, non-strict NPI does. The data where a negation locally licenses NPI do not exhibit any particular variations. Therefore, we show the

data with clause-mate negative licensing in the appendix D for the sake of brevity. The restrictions relating to long-distance licensing are further elaborated below.

In (8) and (9), the NPI occurs in the embedded clause and the negation *-lē-* in the matrix clause depicting long-distance licensing. The constructions show ungrammaticality, since the adjunct clause and the complex NP function as syntactic islands. That is, the negation in the matrix clause cannot license the NPI in the embedded clause.

1. NPIs in adjunct clauses

8a. *rāmu ikkaḍiki **[evari-kī** cepp-i] rā**-lē-**du
Ramu here **whom-with.evenNPI** tell-CPM come**-NEG-**3.SG.NM
'*Having telling anyone, Ramu did not come here.'

8b. *rāmu **[koncem kūḍā** tin-i] paḍu-ko**-lē-**du
Ramu **little even.NPI** eat-CPM sleep-SELFBEN**-NEG-**3.SG.NM
*'Having eaten little also, Ramu did not sleep.'

8c. *rāmu ikkaḍiki **[cilli gavva** unḍi] rā**-lē-**du
Ramu here **single penny.NPI** having come**-NEG-**3.SG.NM
*'Ramu did not come here with single penny.'

2. NPIs in complex NPs

9a. ***[evvarū** vāḍina] katti] nī daggara **lē-**du
who.even.NPI used knife you with be**.NEG-**3.SG.NM
*'The knife which is used by anyone is not with you.'

9b. ***[koncem kūḍā** panḍina panḍu] pullagā **lē-**du
little even.NPI ripen fruit sour-adjl be**.NEG-**3.SG.NM
*'The fruit which ripened at all, is not sour.'

9c. ***[cilli gavva kūḍā** unna] nī-ku] lāṭarī tagala**-le-**du
single penny even.NPI have you-DAT lottery get-PST.**NEG-**3.SG.NM
*'You who have a single penny, did not win a lottery.'

The complement clause as in (10a) shows that a wh-NPI *evaru.u* exhibiting long-distance licensing leads to ungrammaticality as in (10a). The ungrammaticality is because the embedded clause functions as a syntactic island blocking the long-distance licensing. In contrast, q-NPI *konni-kūḍā* and i-NPI *cilli gavva kūḍā* exhibit long-distance licensing as in (10a) and (10b). Due to the possibility of long-distance licensing, q-NPIs and i-NPIs are non-strict.

3. NPIs in complement clauses

10a. ***[[[evaru.u** unn-ā-ru] ani] nēnu anukō**-lē-**du]
who.even.wh-NPI be-PRES-3.PL.H COMP I think-PST.**NEG-**1.SG.NM
'I did not think that there is anybody.'

10b. **[[[konni kūḍā** unṭ-ā-yi] ani] nēnu anukō**-lē-**du]
few-even.q-NPI be-PRES-3.PL.H COMP I think-**NEG.PST**-3.SG.NM
'I did not think that there can be few also.'

10c. [[[nī daggara **cilli gavva kūḍā** un-ṭun-di] ani] anukō**-lē-**du]
you with **single penny even.i-NPI** be-PRES-3.SG.NM COMP think-**NEG.PST** -1.SG.NM
'I did not think that you will be having a single penny also.'

3.5 Occurrence of an NPI in non-negative contexts

The previous studies such as Lahiri (1998), Kumar (2006) and Bhattacharyya (2012) among others discuss the occurrence of NPIs in non-negative contexts such as yes/no question, conditional, imperative, generic, modal of possibility and adversative predicate. The data in the studies mentioned above shows that the morphological composition of NPI is identical in negative as well as in non-negative contexts as in (11) and (12). For example, the NPI in the presence of negation is *any* or *kisii bhii* as in (11) and the NPI in the non-negative context will also be *any* or *kisii bhii* as in (12) (for NPI in non-negative constructions in Hindi see Kumar (2006)). However, Telugu being morphologically rich language, NPIs in non-negative contexts as in (13)–(18) are morphologically different from the one in the presence of overt negation. That is, the NPI is attached with *kūḍā* 'even' in the presence of a sentential negation and in contrast, NPI is attached with *ainā* 'at least' in non-negative context. In the following section, we demonstrate the quantificational restrictions of NPI in non-negative context (NPI-*ainā* hereafter) and the NPI in the presence of overt negation (NPI-*kūḍā* hereafter) and we also describe a correlation at the level of semantic configuration between both the types of NPIs.

11. *Hindi-Urdu*
maiN-ne kisii bhii sTuDeNT ko nahiiN dekh-aa
I-ERG any.NPI student to NEG se-PERF
'I did not see any student.' (Kumar, 2006, p. 109)

12. *Hindi-Urdu: Yes/No Question*
aap-ne kisii bhii sTuDeNT ko dekh-aa (kyaa)
you-ERG some even student to see-PERF what
'Did you see any student?' (Kumar, 2006, p. 111)

13. *Telugu: Yes/No Question*
    ā rūm lo **evar(u)-ainā** unn-ā-r(u)-ā?
    that room in **who-at least** be-PST-3.PL.H-INT
    'Is anybody there in that room?'

14. *Telugu: Conditional*
    okavēla ā gadilō-ki **evarainā** vaste, nēnu nī-ku cept-ā-nu
    if that room-in **who at least** come, I you-to tell-PST-3.SG
    'I will let you know, if anybody comes into the room.'

15. *Telugu: Imperatives*
    ēdainā tinu
    which at least eat-3.SG
    'Eat anything.'

16. *Telugu: Generics*
    yē pilli ainā eluka-ni vēṭāḍu-tun-di
    any cat at least rat-ACC hunt-GEN-3.SG.N
    'Any cat hunts a rat.'

17. *Telugu: Modals of Possibility*
    evvarainā ī tēbl ni etta-galugutā-ru
    who at least this table ACC lift-poss-3.PL.H
    'Anyone can lift this table.'

18. *Telugu: Adversative Predicates*
    nuvvu **ēdainā** ceppāvanṭe nā-ku āscaryangā un-di
    you **who.atleast** tell I-DAT surprising be.PRES-3.SG.NM
    'I am surprised that you told anything to the police.'

Summing up the discussion in this section, Telugu differs from English, where the NPI can occur in subject position in Telugu. In English, Hindi and Telugu, negation licenses multiple NPIs. The complement clauses in Telugu, wh-NPIs are strict NPIs since as they do not allow long-distance licensing. In contrast, q-NPIs & i-NPIs function as non-strict NPIs since they allow long-distance licensing. The NPI in the negative context and in the non-negative contexts show morphological variations. In the following discussion, we claim that NPI-*ainā* and NPI-*kuḍā* are counterparts of a single type of NPI.

4 Semantic description of NPIs: an alternative-based structure

In this section, we discuss the quantificational restrictions of NPIs by adopting analyses of Chierchia (2013) for NPI-*kuḍā* and Nicolae (2012) for NPI-*ainā*. Chierchia's analysis of NPIs provides an answer to the question, "Why the class of NPI licensors is semantically uniform and why NPIs have the shape they do?" The analysis takes place

through the process of feature checking, where the negation is the goal and the NPI is the probe. The probe inherently has a scale with active quantificational alternatives arranged over it, where the alternative to the right entails the alternative to the left. For example, if the scale is <*one, two, three…*>, *two* entails *one.* The scale functions as the uninterpretable feature of the probe, in other words [uNEG]. Negation which is the goal consists of negative features which function as the interpretable features such as the [iNEG]. In line with Lahiri's (1998) analysis of NPIs, the emphatic operator is associated with a low-point element in a scale and this point functions as semantically the strongest alternative. Strongest alternative, in the sense, it functions as threshold, where no other alternatives further entails it. For example, *ēk bhī* 'any.NPI' as in (19) has *ēk,* indicating the numeral 'one', which is a low-point on the active alternative scale. This low-point functions as a strongest alternative, in which case, anything that is entailed within *ēk* are counted. Since the alternatives entailed in *ēk* is zero alternatives, *ēk bhī* in a negative context indicates 'no individuals'.

19. **ēk bhī** ādmī nahĩĩ āyā
    any.NPI man not came
    'No man came.' (Chierchia, 2013, p. 156)

This procedure of selecting the lowest point in the scale and making it semantically strongest is called scale truncation. In other words, the alternative that functions as a threshold is considered the *least* likely alternative and only those alternatives lesser than the threshold are active. The entire process mentioned above occurs only in downward-entailing context.

Nicolae (2012) provides an alternative-based semantic account of PPIs. The analysis lays a connection between a PPI and an NPI. In this analysis, the super-domain alternatives are active in a PPI. That is, the alternatives which entail the emphasized alternative are counted. For example, in the scale <*one, two, three…*> if we suppose that *two* is the emphasized alternative, any numeral entailing *two* is a super-domain. In other words, any numeral greater than or equal to *two,* which are <*two, three…*> are considered to be a part of the super-domain. Such activation of super-domain occurs in upward-entailing context only. In contrast, the sub-domain alternatives are active in the case of NPI. For example, in the scale <*one, two, three…*> if *two* is the emphasized alternative, then the sub-domain would be anything that *two* entails. Hence, the sub-domain includes *one*, since *two* entails *one*. We make a modification to Nicolae's analysis, where we apply the PPI's analysis to NPI-ainā. Further, we do not call NPI-ainā as PPI, since the context is semantically negative, even if it is morphologically/ syntactically affirmative.

The application of Chierchia (2013) and Nicolae (2012) to NPIs in Telugu has the following results. Recall that NPI-*kūḍā* is a counter-part of NPI-*ainā* and the evidence is shown at the morphological level in constructions with non-negative context. We claim

a similar correlation of NPI-*kūḍā* and NPI-*ainā* at semantic level as well. The discussion is elaborated below. The NPI-*kūḍā* and NPI-*ainā* possess an inherent neutral element and we label it as a Polarity Item (PI), which is identical to both NPI-*kūḍā* and NPI-*ainā.* By identical, we mean that the PI has a quantificational scale with active alternatives arranged in a linear, incremental order, where the alternative on the right entails the alternative on the left and one among the alternatives is an emphatic alternative, where it functions as a threshold/mid-point. The quantification on the scale is restricted, consequent to the negation/affirmation influencing the PI. Note that the particle *kūḍā* or *ainā* realizes only following the feature checking depending on what type of restriction exists on the scale. In the case of an NPI-*kūḍā*, due to the influence of the overt negation, the sub-domain alternatives remain active and the super-domain alternatives are cancelled. In the case of NPI-*ainā*, due to the influence of the overt affirmation, the super-domain alternatives are active and the sub-domain alternatives are cancelled. The NPI-*kūḍā* and NPI-*ainā* are related at semantic level also. That is, the structure of alternatives that we demonstrated for NPI-*ainā,* consistently occurs in all the non-negative contexts as in (15)–(18).

Summing up the discussion above, we illustrated the quantification structure of NPIs in Telugu based on the alternative-based semantic analysis by adopting Chierchia (2013) and Nicolae (2012). The NPIs inherently possess a neutral element which has a scale with active alternatives. Due to the influence of a negative or affirmative features, the neutral element undergoes cancellation of certain active alternatives depending on the type of NPI. Similar to the correlation at the morphological level, NPI-*kūḍā* shows a correlation with NPI-*ainā* at their semantic level also, where the NPI-*ainā* consistently shows active super-domain alternatives in all the non-negative contexts. NPI-*kūḍā* has active sub-domain alternatives, NPI-*ainā* has active super-domain alternatives. We discuss the feature checking of NPIs along with their licensing conditions in the following section.

5 Licensing conditions of an NPI

In this discussion, we demonstrate an analysis which is a combination of syntactic and semantic operations. We compare previous studies such as Mahajan (1990), Chomsky (1995), Kumar (2006) and we conclude that Kumar's analysis of c-commanding best suits NPIs in Telugu. In addition to Kumar's analysis, we adopt the Chomsky's (2000) feature checking similar to operation Agree and also Chierchia's quantificational restriction. The analysis is elaborated below.

Mahajan (1990) states that the negative licenses the NPI at the level of LF, where the negative moves to a position higher than the NPI, adjoining the finite IP so that the negative c-commands the NPI. Mahajan's analysis encompasses the fact that the

negation c-commands the NPI. However, the analysis may not be suitable for Telugu, particularly for construction depicting long-distance licensing as shown in (20).

20. **lē**$_i$ **[evarinī]**$_i$ [$_{S1}$[$_{S2}$ rāmu$_j$ ikkaḍa **t**$_j$ cūḍa-**t**$_i$-du $_{S2}$] ani] sīta$_i$
neg **who even.NPI** Ramu here see-**NEG**-3.SG.NM] comp Sita
cepp-in-di $_{S1}$]
tell-PST-3.SG.NM]
'Sita said that Ramu did not see anyone here.'

In (20), the NPI and the negative base-generate within the embedded clause and the NPI is scrambled out of its position. Under Mahajan's analysis the negation moves and it adjoins the finite IP at LF in a way that the negative c-commands the NPI. However, the adoption of Mahajan's analysis has the following problems: a. there is no limit to the number of heads moving. b. The movement of the negation is long-distance i.e. it moves from the embedded clause to the left of the finite IP. Such movement is a violation of head-movement constraint. c. The negation cannot move above the adjoined NP, since the adjunction functions a barrier for movement. d. Before the movement, the negative only negates the embedded clause. However, after the movement of the negative, to form an adjunct of a finite IP, the negative negates the entire sentence, that is, the embedded and the matrix clause. As a result, it is a violation of the structure preserving principle, since there is a change in the scope at PF and LF.

Chomsky's (1995) reconstruction states that the NPI is reconstructed at a lower position, so that the negative c-commands the NPI. However, the analysis may not be suitable for analysing NPIs in Telugu.

21. *Surface Structure*
[[sīta$_i$ tīsin-a yē fōṭō kūḍā]$_j$ tanaki$_i$ nacc-alēdu ani] rāmu
Sita take-ADJL Photo even.NPI her like-not COMP Ramu
cepp-ā-ḍu t$_j$
tell-PST-3.SG.M
'Ramu said that she does not like any photograph that Sita took.'

22. *Logical Form*
*e$_j$ tanaki$_i$ nacc-alēdu ani rāmu cepp-ā-ḍu [sīta$_i$ tīsin-a
her like-not COMP Ramu tell-PST-3.SG.M Sita take-ADJL
yē fōṭō kūḍā]$_j$
Photo even.NPI
'Ramu said that she does not like any photograph that Sita took.'

In (21) and (22), we find that *sīta tīsin-a fōṭō* 'the photograph which is taken by Sita' is scrambled to sentence initial position, in a way that *sīta,* a referential expression c-commands *tanaki* 'to her', a pronoun. However, when the scrambled element is reconstructed at the level of LF, then *tanaki* wrongly c-commands *sīta.* We consider it 'wrongly c-commanding' in (22), since *sīta,* a referential expression is supposed to be

free everywhere. Further, not all NPIs occur in a scrambled position. This method of reconstruction may not be applicable for NPIs which are not scrambled.

Kumar's (2006) analysis of Hindi can be applied to NPIs in Telugu, since the problems that arose due to the application of the analyses mentioned above, do not arise in this analysis. Further, the analysis is suitable to account for strict NPIs, non-strict NPIs, multiple NPIs and NPI in subject position in Telugu and Hindi as discussed in sections 3. Kumar's analysis states that the negative licenses the NPI at Deep Structure prior to movement at Surface Structure. Since the structure of Telugu is similar to Hindi, we adopt the analysis to account for NPIs in Telugu, where the negative licenses the NPIs at the base-generated position itself. Hence, even if the NPIs move out of their positions for achieving further operations such as Case and Agreement, they do not lose the negativity of the NPIs.

In order to license an NPI, the negative c-commands an NPI-*kūḍā* and an affirmative c-commands a NPI-*ainā*. Hoeksema (2000) provides evidence that more than c-commanding, it is the scope of negation and negative operators that license an NPI. However, we claim that the negative obligatorily c-commands the NPI at the level of LF. In order to prove our claim, we first illustrate how the strategy of c-commanding operates, when the NPI is the subject of a clause with a local negative licensor. Further, an illustration of an ungrammatical construction is provided, where the NPI occurs outside the c-commanding domain of the negative at the level of LF. The illustration is elaborated with the help of complex sentences below.

The construction in (23) has a structure depicted in (24), where the negative c-commands the NPI, prior movement to subject position. The operations do not lead to ungrammaticality.

23. [akkaḍa **okkaḍu kūḍā** **lē-**ḍu] ani anukunn-ā-nu
there **one person even.NPI** **NEG-**3.SG.M COMP think-PST-1.SG
'I thought that there is nobody.'

24. *Structure of example 23:*

LF: [$_{matrix.TP}$ - - - [$_{embedded.TP}$ e$_i$T [$_{NegP}$ **Neg** [$_{vP}$ **NPI.Subj$_i$** v [$_{VP}$ V Obj]]]]]

c-commanding

PF: [$_{matrix.TP}$ - - - [$_{embedded.TP}$ **NPI.Subj$_i$** T [$_{NegP}$ **Neg** [$_{vP}$ t$_i$v [$_{VP}$ V Obj]]]]]

no c-commanding

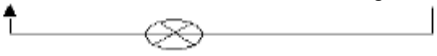
Now, we demonstrate how the occurrence of an NPI outside the c-commanding domain of negation at the level of LF leads to ungrammaticality. The construction in (25) has a structure depicted in (26), where the negative does not c-command the NPI, both at LF and PF. If we compare the LF in (24) and the LF in (26), we notice that the c-commanding strategy exists in (23) which is the structure of a grammatical sentence.

Based on the correspondence between grammaticality and c-commanding in (23), we claim that the strategy of c-commanding is a necessary condition for the sentence to be grammatical. We further state that c-commanding also exists in licensing an NPI-*ainā*, where an affirmative licenses it, at the base-generated positions prior to movement for further operations.

25. *[akkaḍa **lē-**ḍu] ani **okkaḍu kūḍā** anukunn-ā-nu.
there **NEG-**3.SG.M COMP **anyone.NPI** think-PST-1.SG
'I thought that there is nobody.'

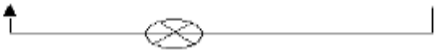
26. *Structure of example 25:*

LF: $[_{matrix.TP}\ e_i\ T\ [_{vP}\ \mathbf{NPI.Subj_i}\ [_{VP}\ V\ [_{embedded.TP}$ - - - $[_{NegP}\ \mathbf{Neg}\ [_{vP}$ - - -]]]]]]



no c-commanding

PF: $[_{matrix.TP}\mathbf{NPI.Subj_i}\ T\ [_{VP}\ V\ [_{embedded.TP}$ - - - $[_{NegP}\ \mathbf{Neg}\ [_{vP}$ - - -]]]]]]



no c-commanding

However, c-commanding is not the only condition that is required to license NPI. If we adopt c-commanding as the only condition, the negation wrongly licenses every element that occurs in the c-commanding domain. In addition to c-commanding, the NPIs must have two more properties inherently such as, an active alternative scale and *ainā* 'epistemic *at least*'. These two properties play a role in feature interaction with their respective licensors. Feature checking takes place between negation & NPI-*kūḍā* and affirmation & NPI-*ainā*. Prior to the feature checking the NPI remains a neutral item as discussed in section 4. Below, we provide the feature composition that leads to the realization of different types of NPIs. In the configuration mentioned below, the properties on the left hand side together form either NPI-*kūḍā* or the NPI-*ainā* mentioned on the right hand side. The licensor identifies that an element is a NPI, by checking the active alternative scale inherent to the NPI. Further, the NPI takes the morphological realization based on the function of the particle, which is inherently present in it. A schematic representation of feature checking is provided in (27).

27. Before: $[_{NegP}\ Neg_{[iNeg]}\ [_{vP}\ NPI.Subj_{[u\text{-}Neg]}\ v\ [_{VP}\ V\ NPI.Obj_{[u\text{-}Neg]}]]]$
After: $[_{NegP}\ Neg_{[iNeg]}\ [_{vP}\ NPI.Subj_{[u\text{-}Neg]}\ v\ [_{VP}\ V\ NPI.Obj_{[u\text{-}Neg]}]]]$

In the case of NPI-*kūḍā,* the PI with the scale and the feature of emphatic 'even', checks its uninterpretable features with its goal, which has interpretable negative features, resulting in NPI-*kūḍā* as in (a). In the case of NPI-*ainā,* the PI with the scale and feature of epistemic *at least*, checks its uninterpretable features with its goal, which has interpretable affirmative features as in (b).

a. PI + negation + scale + inclusiveness feature → sub-domain alternatives of the threshold are active → NPI-*kūḍā.*

b. PI + affirmation + scale + epistemic feature → super-domain alternatives greater than or equals to the threshold are active → PPI-*ainā*.

Summing up the discussion in this section, adopting Kumar (2006), we state that the negative/affirmative licenses the NPI vP-internally at the level of LF. Consequent to their occurrence in a c-commanding domain of the licensor, the polarity-sensitive items undergo feature checking, where the negative/affirmative function as the goal and the NPI as the probe. The probe is a neutral item with uninterpretable features such as the active alternative scale, along with one of the functions such as epistemic *at least,* dubitative or emphatic properties. Further, a goal is either an affirmative licensor or a negative licensor. In the process of feature checking, the probe checks its uninterpretable features with the interpretable features of the corresponding goal. As a result, the PI which is a neutral item morphologically realizes into NPI-*kūḍā* or NPI-*ainā*.

6 Conclusion

In this paper, we described in detail, the distribution of NPIs followed by the analysis of licensing conditions. The distribution of NPIs covers the following. There are three types of elements that occur as the root of the NPIs: wh-elements (wh-NPI), quantifiers (q-NPI) and idiomatic expressions (i-NPI). When these elements occur in complement clauses, wh-NPIs function as strict NPIs, since they disallow long-distance licensing. In contrast, q-NPIs and i-NPIs function as non-strict NPIs, since they allow long-distance licensing. The NPIs occur in subject position, unlike English. Further, multiple NPIs can occur in a single clause. We notice that the NPI in the non-negative contexts possess a morphological structure different from the NPI in the presence of overt negation. It is NPI-*ainā* in the non-negative contexts which is the counterpart of NPI-*kūḍā* in the negative contexts. The evidence for such correlation between NPI-*ainā* and NPI-*kūḍā* is based on the consistent occurrence of *ainā* 'at least' attached to the NPI in every non-negative construction. Further, a similar kind of consistency is noticed at the semantic level, where NPI-*ainā* in every non-negative context depict activation of super-domain alternatives. Based on the distribution of the NPIs mentioned above, we illustrated that the negative/affirmative licenses NPI-*kūḍā/ainā*, when the NPI base-generate at vP-internal positions at LF, prior any type of movement for Case/Agreement. Parallel to c-commanding, the polarity-sensitive items undergo feature checking. The negation/affirmation is the goal with negative/affirmative interpretable features. The polarity-sensitive item prior to feature checking is a neutral item (PI), a probe, which has uninterpretable features: active alternative scale and one of the functions such as emphatic *even* or epistemic *at least*. This scale is the main

feature that distinguishes the polarity-sensitive item from other element in the c-commanding domain. It is the function of the particle, inherently located within the polarity item, which is also responsible for the type of NPI. In feature checking, the probe checks its uninterpretable features with the interpretable features of the goal, as a result, the NPI is realized into NPI-*kūḍā* or NPI-*ainā*.

References

Balusu, R., Gurujegan M., & Rajamathangi S. (April 5 - 8 2016). Bagel Problem Items in Telugu and Tamil. [Presention]. *GLOW 2016 - 39th Generative Linguistics in the Old World.* Georg-August-Universitat Gottingen.

Bhattacharyya, A. (2012). *Polarity sensitive Any in Bengali*. Master's thesis. Retrieved from http://triceratops.brynmawr.edu/dspace/handle/10066/10701.

Chierchia, G. (2013). *Logic in grammar: Polarity, free choice, and intervention*. OUP Oxford.

Chomsky, N. (1991). Some Notes on the Economy of Derivation and Representation. In R. Freidin (Ed.), *Principles and Parameters in Comparative Grammar* (pp. 417-454). Cambridge: MIT Press.

Chomsky, N. (1995). *The minimalist program*. Cambridge. MA: MIT press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels & J. Uriagereka (Eds.), *Step by Step: Minimalist Essays in Honour of Howard Lasnik* (pp. 89-155). Cambridge: MIT Press.

Gajewski, J. (2007). Neg-raising and polarity. *Linguistics and Philosophy*, *30*(3), 289-328.

Hoeksema, J. (2000). Negative polarity items: Triggering, scope and c-command. *Negation and Polarity*, 115-146.

Krishnamurti, B., & Gwynn, J. P. L. (1985). *A grammar of modern Telugu*. Oxford University Press.

Kumar, R. (2006). *Negation and licensing of negative polarity items in Hindi syntax*. Taylor & Francis.

Kuno, S., & Whitman, J. (2004). Licensing of multiple negative polarity items. In S. Kuno, Y.-K. Kim-Renaud & J. Whitman (2004). *Studies in Korean syntax and semantics* (pp. 207-228). Seoul: International Circle of Korean Linguistics.

Ladusaw, W. (1979). *Polarity sensitivity as inherent scope relations*. Unpublished PhD thesis, University of Texas at Austin, Austin.

Ladusaw, W. A. (1983). Logical form and conditions on grammaticality. *Linguistics and Philosophy, 6*(3), 373-392.

Lahiri, U. (1998). Focus and negative polarity in Hindi. *Natural Language Semantics*, *6*(1), 57-123.

Laka, I. (1989). Constraints on sentence negation. In I. Laka & A. K. Mahajan (Eds.) *Functional heads and clause structure* (pp. 199-216). Cambridge: MIT Press.

Laka, I. (2016). *On Syntax of Negation*. Routledge.

Lasnik, H. (1972). *Analysis of negation in English*. (Unpublished doctoral dissertation), MIT.

Lee, Y. S., & Horn, L. (1994). *Any as indefinite plus Even*. Manuscript. New Haven: Yale University.

Linebarger, M. C. (1980). *The grammar of negative polarity*. (Unpublished doctoral dissertation). MIT.

Mahajan, A. K. (1990). LF conditions on negative polarity item licensing. *Lingua, 80*, 333-348.

Nicolae, A. (2012). Positive Polarity Items: an alternative-based account. In *Proceedings of Sinn und Bedeutung*. Vol. 12.

Pollock, J. Y. (1989). Verb movement Universal Grammar and the structure of IP. *Linguistic Inquiry, 20*, 365-424.

Progovac, L. (1991). Polarity in Serbo-Croatian: Anaphoric NPIs and pronominal PPIs. *Linguistic inquiry, 22*(3), 567-572.

Progovac, L. (1994). *Negative and positive polarity: A binding approach*. Cambridge: Cambridge University Press.

Ramachandran, S. (1991). *Negativity in Tamil: Untying the undefinable not*. University of Ottawa.

Zwarts, F. (1998). Three types of polarity. In E. H. F. Hamm (Ed.), *Plural quantification* (pp. 177-238). Dordrecht: Kluwer.

APPENDIX A

A list of quantifiers

The following list shows quantifiers attached with particle indicating 'even'. The non-occurrence of a particle with the quantifier is depicted with '*'.

Table A1: Universal Quantifiers

Telugu	Gloss
andaru-u; *andaru-kuḍā/*antamandi.i; antamandi-kuḍā	that many people even [+human]
anni.i/anni-kuḍā	that many even [-animate][+countable']
*anta.a/anta-kuḍā	that much even [-animate][-countable']

Table A2: Existential quantifiers

Telugu	Gloss
*kondaru.u; kondaru kuḍā/kontamandi.i; kontamandi kuḍā	few people even [+human]
konni.i/konni kuḍā	few things even [-animate][+countable']
*koncemu.u; koncem kuḍā/*konta.a; konta kuḍā	little even [-animate][-countable']

APPENDIX B

A list of numerals

The following list shows numerals attached with particle indicating 'even'. The non-occurrence of a particle with the quantifier is depicted with '*'.

Table B1: Cardinal numerals with [+human] feature

Telugu	Gloss
okkaḍu.u; okkaḍu kuḍā	one person even
iddaru.u; iddaru kuḍā	two persons even

Table B2: Cardinal numerals with [-human] feature

Telugu	Gloss
okaṭi.i; okaṭi kuḍā	one even
reṇḍu.u; reṇḍu kuḍā	two even

Table B3: Ordinal numerals

Telugu	Gloss
modaṭidi.i/ modaṭidi kuḍā	first one even
reṇḍavadi.i/ reṇḍavadi kuḍā	second one even

APPENDIX C

A lists of wh-entities

The following list shows wh-elements attached with particle indicating 'even'. The non-occurrence of a particle with the quantifier is depicted with '*'.

Table C1: wh-entities

Telugu	Gloss
enduku.u; enduku kuḍā/dēniki.i; dēnikikuḍā	why even
ēmi.i; *ēmi kuḍā	what even
ekkaḍa.a; ekkaḍa kuḍā	where even
eppuḍu.u; eppuḍu kuḍā	when even
*ela.a; *elā kūḍā	how even
evaru.u; *evaru kūḍā	who even
ēvi.i/ *ēvi kuḍā	which ones even
ēdi.i/ ēdi kuḍā	which one even

APPENDIX D

The negation licenses the NPIs locally. All the three types of NPIs can be licensed by a local negative licensor. The NPIs do not show any particular variations in the structure when they are locally licensed. NPIs depicting long distance licensing in similar type of complex sentences show variations in the licensing conditions which is elaborated in section (3.4).

1. NPIs in adjunct clauses

i a. rāmu ikkaḍiki **[evari-kī** cepp-**akunḍā**] vacc-ā-ḍu
Ramu here **anybody-with.NPI** tell-**without.Neg** come-PST-3.SG.M
'Ramu came here without telling anyone.'

i b. rāmu **[koncem kūḍā** tin-**akunḍā]** paḍu-konn-ā-ḍu
Ramu **little even.NPI** eat-**without.Neg** sleep-selfben-PST-3.SG.M
'Ramu slept without eating anything.'

i c. rāmu ikkaḍiki **[cilli gavva lēkunḍā]** vacc-ā-ḍu
Ramu here **single penny.NPI without.Neg** come-PST-3.SG.M
'Ramu came here without a single penny.'

2. NPIs in Complex NP

ii a. **[evvarū** vāḍ-**ani]** katti] nī daggara un-di
anybody.NPI use-**neg** knife you with be-3.SG.NM
'The knife which is not used by anyone, is with you.'

ii b. **[koncem kūḍā** paṇ-**ani** paṇḍu] pullagā un-di
little even.NPI ripe-**neg** fruit sour-adjl be-3.SG.NM
'The fruit which did not ripen little bit also, is sour.'

ii c. **[cilli gavva kūḍā leni]** nī-ku] lāṭarī tagil-in-di
single penny even.NPI not have you-dat lottery get/win-PST-3.SG.NM
'You who do not have even a single penny, won a lottery.'

3. NPIs in complement clauses

iii a. **[[[evarū lē-**ru] ani] nēnu anu-konn-ā-nu]
anybody.NPI be.PRES.**NEG-**3.PL.h comp I think-selfben-PST-1.SG
'I thought that there is nobody.'

iii b. **[[[konni-kūḍā lē-**vu] ani] nēnu anu-konn-ā-nu]
few-even.NPI be.PRES.**NEG-**3.PL.NH comp I think-selfben-PST-1.SG
'I thought that there cannot be few also.'

iii c. [[[nī daggara **cilli gavva kūḍā lē-**du] ani]
you with **single penny even.NPI** be.**NEG.**PRES-3.PL.NH comp
anu-konn-ā-nu]
think-selfben-PST-1.SG
'I thought that you don't have a single penny also.'

Abbreviations

1	: first person
3/iii	: third person
ACC	: accusative
ADJL	: adjectivaliser
COMP	: complementizer
CPM	: conjunctive participle marker
DAT	: dative
DUB	: dubitative
EMPH	: emphatic
ERG	: ergative
F	: feminine
FUT	: future
GEN	: genitive
H	: human
INDEF	: indefinite
INF	: infinitive
LOC	: locative
M	: masculine
N	: neuter
NEG	: negative
NH	: non-human
NM	: non-masculine
PERF	: perfective
PL	: plural
PRES	: present
PST	: past
QP	: quotative particle
SELFBEN	: self-benefactive
SG	: singular

INTEGRATION FUNCTIONS OF TOPIC CHAINS IN CHINESE DISCOURSE

Kun SUN
University of Tuebingen, Germany
Zhejiang University, China
kun.sun@uni-tuebingen.de; sharpksun@hotmail.com

Abstract

Topic chain, one of the essential organization devices in Chinese discourse, is highlighted by the use of many co-referential zero forms. Although topic chain plays an important role in organizing discourse, few attempts have been made to explore how topic chain forms an integrated and meaningful unit, and how it facilitates discourse organization through the so-called "integration functions". This study, based on a comprehensive review of topic chain studies, re-examines the core characteristics of the topic chain. Later on, integration functions of the topic chain are analysed at internal and external levels. Topic chain itself can manage its internally different clauses to form a cohesive, meaningful and unified unit. At this stage, this paper clearly demonstrates why so much information within a topic chain assembles in a compact structure. At the discourse level, one topic chain can associate with another topic chains or non-chain constructions to establish textual coherence. Making the use of zero anaphora, co-reference, cognitive orders, and other non-morph-syntactic devices, topic chain can combine different discourse units together to construct Chinese discourse. The study provides a systematic and well-developed account of the integration functions for the Chinese topic chain, which plays a significant role in understanding the nature of a topic chain as well as in understanding on how discourse coherence establishes in Chinese.

Keywords: co-referential topic; connection of clauses; unified unit; textual coherence; compactness; meronymy

Povzetek

Tematsko verigo, eno izmed osnovnih organizacijskih sredstev v kitajskem diskurzu, zaznamuje uporaba številnih soreferenčnih ničelnih oblik. Navkljub dejstvu, da tematska veriga igra pomembno vlogo pri organizaciji kitajskega diskurza, je le malo raziskav na temo, kako le-ta oblikuje celostno in smiselno enoto ter kako pripomore k lažji organizaciji diskurza preko t.i. itegracijskih funkcij. Študija, ki temelji na celovitem pregledu raziskav o tematskih verigah, ponovno preverja glavne značilnosti tematske verige. Hkrati analizira integracijske funkcija na zunanjih in notranjih ravneh verige. Tematska veriga lahko upravlja

Acta Linguistica Asiatica, 9(1), 2019.
ISSN: 2232-3317, http://revije.ff.uni-lj.si/ala/
DOI: 10.4312/ala.9.1.29-57

s stavki, ki so notranje različni, in iz njih oblikuje oblikuje smiselno povezano, celostno enoto. V tem delu študija jasno prikaže, zakaj se tako veliko informacij zbere v kompaktni strukturi. Na nivoju diskurza se ena tematska veriga poveže z drugo tematsko verigo ali kakšno drugo strukturo in tako se vzpostavi besedilna skladnost. Z uporabo ničelne anafore, soreferenčnosti, kognitivnega zaporedja in drugih pripomočkov, ki niso morfološki ali sintaktični, je tematska veriga zmožna povezati različne diskurzne enote v kitajski diskurz. Študija ponuja sistematičen in dobro razvit pregled integracijskih funkcij za kitajsko tematsko verigo, kar močno prispeva k razumevanju narave tematskih verig kot tudi k razumevanju načina vzpostavljanja diskurzne skladnosti v kitajščini.

Ključne besede: soreferenčna tema; povezava med stavki; poenotena celota; besedilna skladnost; kompaktnost; meronimija

1 Introduction

The Chinese language is known as a discourse-oriented language (Tsao, 1979, 1990; Chu, 1998; Li, 2005). Meanwhile, Chinese is regarded as a topic-prominent language (Li & Thompson, 1976, pp. 457-461). Topic chain is the intersection of the two salient characteristics, as well as the bridge between syntax and discourse. It is characterized by the use of co-referential zero forms in successive different clauses, and is regarded as one of the typical landmark characteristics of the Chinese language.

In comparison with English, topic chain in Chinese is a compact construction stored with large qualities of information, and readers of other languages will wonder why so much information is assembled into a compact and short structure, and how the structure works effectively.

Most studies on topic chains tend to explore how a topic and its co-referential zeros are produced within a topic chain, but ignore how topic chains form a larger propositional meaningful unit or establish textual coherence. An innovative term, “integration function”, is used here to specify how a topic chain works to construct small units into a meaningful coherent unification and create textual coherence in Chinese. The term includes two layers, i.e. a topic chain is able to construct its inner unification as an integrated and meaningful unit, and a topic chain can establish textual coherence through cooperating other topic chains or other constructions. Being different from textual coherence, this term “integration function” is characteristic of its form (a structural configuration) and function (coherence establishment) for a topic chain. The purpose of the study is to probe into integration functions of a topic chain from a textual perspective.

The present study first reviews literature on the Chinese topic chain by presenting their linear features. Following this, the paper focuses on integration functions of the topic chain, which is the theme of the study.

2 A brief review of studies on Chinese topic chain

2.1 Definitions of topic chain

The term “topic chain” was first put forward by Dixon (1972, p. 71), who investigated an Australian aboriginal language Dyirbal. Tsao (1979) first defined the concept of topic chain in Chinese, which is different from that of Dixon. After him, many linguists worked on Chinese in this field.

Tsao (1979, p. vii) gives his definition of a topic chain for Chinese as, “a topic chain, a stretch of actual discourse composed of one –and often more than one – clause, headed by a topic which serves as a common link among all clauses, that actually functions as a discourse unit in Chinese”. He gives an example: (Tsao, 1979, p. 38)

(1) **那 棵 树，** 花 小 ø 叶子 大，ø 很 难 看，
nà kē shù huā xiǎo ø yèzi dà ø hěn nán kàn
That M tree flower small, (tree) leaves big, (tree) very plain looking,

所以 我 没 买。
suǒyǐ ø wǒ méi mǎi
so (tree) I NOT buy.

According to Tsao, a topic chain can occur in a simple sentence, where he defines two types: a simple-sentence topic chain and a complex-sentence topic chain. Tsao’s definition has been accepted and cited by Li (1985, p. 138), Shi (1992, p. 15), and Li (1995, p. 32).

Li & Thompson (1981, p. 659) define a topic chain as follows.

> One common situation in which noun phrases are unspecified is the topic chain, where a referent is referred to it in the first clause, and then there follow several more clauses talking about the same referent but not overtly mentioning that referent.”

Chu (1998, p. 324) offers a more inclusive definition stating that “a topic chain is a set of clauses linked by a topic in the form of ZA (zero anaphora)”. His typical example is shown as (Chu, 1998, p. 335):

(2) 一天， 趁 他 洗澡，**我** 边 去 检查
yì tiān, ø chèn tā xízǎo, **wǒ** biàn qù jiǎnchá
One-day, (I) take-advantage-of he bathe, I at-once go search

他 的 衣服， 翻 了 上衣 的 每个 口袋，
tā de yīfu, ø fān le shàngyī de měi gè kǒudɑi,
his DE clothes, (I) turn-over- PFV top-clothes’ DE every-M pocket,

又	去	翻	裤子	的	口袋。
yòu	qù	fān	kùzi	de	kǒudɑi
further	go	turn-over	slack	DE	pocket

Chu's definition is strikingly similar to Li and Thompson's, but he introduces another term, the so-called "zero anaphora**",** which had not been explicitly mentioned before.

Although Li (2005) does not present a clear definition, her underlying assumption about a topic chain is basically in accordance with Chu's, where zero anaphora is regarded as the core characteristic. Further, Li (2005, p. 56) holds that the overt topic of a topic chain does not have to occur in the initial clause of a chain, and that some zero NPs can be linked to an overt topic not in the immediate sentence, but in the previous sentence or even previous paragraph, instead. Therefore, Li suggests more forms of a topic chain, such as multi-sentence topic chain; multi-paragraph topic chain; discontinuous topic chain; modifier topic chain.

These definitions all implicitly agree that each clause in a topic chain shares the same topic, but as for the performance requirements of these topics, they display substantial differences. Tsao has no description for the performance of a topic in each clause, while Li and Thompson (1981), Chu (1998) and Li (2005) all are in favor of using zero forms within a topic chain. However, they differ slightly from each other: Li and Thompson think that the overt topic should be located in the initial-sentence position; Chu believes that the topic does not have to appear in the initial position. The clause (the topic clause) with the topic mentioned firstly often occurs at the beginning of the topic chain, but it also possibly occurs in middle of or at the end of the chain (Li, 2004).

By comparison, Li has a more inclusive analysis of zero form performance in a topic chain, and summarizes three features of this phenomenon (Li, 2005, pp. 54-55):

a) a topic is overtly mentioned in the first clause;
b) a topic is anaphorically referred to by a zero form in the subsequent clause(s); and
c) a topic chain is basically a sentence.

What is a topic chain? Generally, the nature of a topic chain can be explored from several aspects such as including its grammatical category, behaviour and function, just to name a few. The perspective of its category is one such aspect. So far, most studies have failed to explicitly delineate its category of a topic chain except for Shi (1989), who considered a topic chain as a syntactic category. Despite of the implicit descriptions, most studies basically agree that a topic chain is a discourse category though it can play an important role in syntax. The discussions on the definitions of the topic chain in previous studies show that making of a topic chain needs the same topic and zero forms, but there nevertheless remain some problems that no previous study seems to deal with. Is a topic chain a unified and meaningful unit, or just a loosely structured discourse? If it is an integrated unit, how do its inner components work together?

With regards to functions of a topic chain, studies agree that a topic chain is an important device in organizing Chinese discourse, but they have seldom demonstrated how topic chains help build up discourse coherence in detail.

2.2 Behaviour of the co-referential topic within a topic chain

The format of the topic chain can be roughly shown as in the diagram below: a shared topic (T) is followed by a series of different comment constituents (C1, C2, Cn), and a co-referential zero form with a topic is situated in the initial position of each comment clause. Each comment constituent and invisible zero form can be considered to be a clause within a topic chain, such as ø C2, and ø Cn, and is therefore called a comment clause. T-C1 is called a topic clause.

T——C_1……（ø）C_2……（ø）C_3……（ø）C_n

There raises a question on whether a visible co-referential form in the position of C2, C3 or Cn symbolizes that the topic chain will come to the end.

Actually, a topic chain is strung by a topic clause and several comment clauses sharing a co-referential topic, and zero forms within a topic chain are maintained by a co-reference. Co-reference is a topic mechanism which controls each comment clause, so it is assumed that a topic can work to delete each co-referential topic in a clause. In this way, zero forms can be considered to be the result of a deletion run by topic. Therefore, zero forms should not be regarded as the only criteria to judge the boundary of a topic chain. Instead, a co-referential topic is the underlying criteria which should be used to determine the boundary of a topic chain. In most cases, a co-referential topic occurs as a zero form, however, it can also occur visibly (pronominal or nominal forms).

These previous studies show that the following two qualities can constitute a topic chain: zero anaphors and an overt topic (occurring once or invisible). It is important to note that on the surface, the two qualities are able to constitute a topic chain, but besides this, there exist two hidden qualities which ensure the build-up of a topic chain: a co-referential topic and zero forms co-referential with a topic. The co-referential topic is the underlying mechanism to control zero forms instead of zero forms themselves.

Table 1: The possible qualities of a topic chain

underlying		**superficial**	
the co-referential topic	zero forms co-referential with the topic	zero anaphors	an overt topic (occurring once or invisible)?

Li and Thompson (1981), Chu (1998) insist that the referent (co-referential with a topic) should not overtly be mentioned, and an overt form co-referential with the topic will only occur once, or not occur at all, but they have different ideas whether the overt

topic should be located in the initial-sentence position. In term of their descriptions, a topic chain is tolerant of an overt topic, indicating that the occurrence of an over topic has no impact on a topic chain which is propped by the underlying qualities (a co-referential topic and zero forms co-referential with a topic). Suppose an overt topic or its co-referential form occurs twice or more in a cluster of clauses, the two underlying qualities are maintained in a cluster of clauses. In other words, when zero forms co-referential with the topic prevail in the cluster in spite of the occurrence of two or more overt topics, all qualities do not violate essential qualities of a topic chain, which carries an over topic or no topic at all. The case of two overt topics is identical with the situation where an overt topic occurs. If an overt topic is seen as one part of a topic chain, two (or more) overt topics are equally treated as parts of a topic chain under the condition of zero forms co-referential with a topic being maintained. However, there is an extreme case where all co-referential topics are not manifested by zero forms, and the cluster of clauses in this case is not seen as a topic chain because the frequent occurrence of overt topics results in no zero forms, completely damaging one underlying quality of making a topic chain.

So we claim that when the occurrence of overt topics has no impact on underlying qualities, the cluster is still treated as a topic chain. Since the occurrence of an overt topic has no direct relation with the underlying qualities, it cannot be treated a necessary quality in making a topic chain despite of the fact that an overt topic often occurs in it once. In such a sense, the requirement of an overt topic just occurring once is neither rational nor justified. When an overt form co-referential with the topic occurs twice or more with zero forms co-referential with the topic being prevalent in a cluster of clauses, the cluster is still a topic chain. Table 2 is a hierarchy of these qualities where ">" indicates that the previous is more important than the latter.

Table 2: A hierarchy of the qualities for a topic chain

the co-referential topic	>	zero forms(anaphors) co-referential with the topic	>	zero anaphors	>	an overt topic (visible or invisible)

It is concluded that the clause where the overt co-referential forms with a topic occur should be considered as one part of a topic chain. A topic chain in a real context can be shown as the following diagram:

T——C_1……（ø/ T'）C_2……（ø）C_3……（ø）C_n

'ø/ T' signifies that zero form, or overt co-referential form, can occur in this position. Compared to the first diagram, this diagram shows that the clause, containing an overt form co-referential with a topic, can be seen as a part of a topic chain.

3 Integration functions of a topic chain

Two clauses in English are combined through two different means: subordination and coordination. Sentences or clauses are coherently integrated through different cohesion devices, such as reference, ellipsis, lexical conjunction (Halliday & Hasan, 1976). Also, more connective devices, as specified by Quirk et al. (1985, pp. 1437-1487), which are more elaborate than those given by Halliday and Hasan, such as pragmatic and semantic implications, intonation, punctuation, and information processing, were identified.

Having special faculties, a topic chain in Chinese can work to connect small units into discourse by making use of their own mechanisms. A topic chain in Chinese performs its integration functions interiorly and exteriorly: firstly, a topic chain itself can manage its different internal clauses to form a cohesive, meaningful and unified unit, without relying on many grammatical devices or markings. Secondly, at the discourse level, one topic chain can cooperate with other topic chains or non-chain constructions to establish textual coherence. Sections 3.1 and 3.2 will discuss its inner and external functions, for which the term "integration function" is used in this study, separately.

3.1 The inner unification of a topic chain

A topic chain is considered to have a topic followed by several comment clauses, as shown in the diagram T——C_1……（ø）C_2……（ø）C_3……（ø/T'）C_n. Within a topic chain, a topic has the power to control and manage its comment clauses, and meanwhile comment clauses are linked coherently, dependent on some mechanism.

3.1.1 Topic's control on comment clauses

Firstly, a topic can control each comment clause through its co-referential mechanism. It is assumed that the topic can work to delete each co-referential topic in a clause. In this way, zero forms can be considered to the result from a deletion performed by the topic, so its co-referential form is encoded as a zero form.

Secondly, a topic chain permits other syntactic constructions to enter, and the inserted construction (embedment) exerts almost no influence on the topic's effective control of its subsequent comment clauses, as illustrated below.

(3) [i] (a) **他** 却 带 少年 喜事 得来 的 脚 疯痛，[…]
tā què dài shàonián xǐshì délái de jiǎo fēngtòng,
he yet bring youngster happy-event attain MOD feet pain,

(b) 买 了 一 条 六桨 白 木 船，
ø mǎi le yì tiáo liù jiǎng bái mù
(he) buy PFV a M six-oared white wooden boat,

(c) 租 给 一 个 穷 船 主，[……]
ø zū gěi yí gè qióng chuán zhǔ，
(he) rent give a M poor boat

(d) 气运 好，
ø qì yùn hǎo
(He) luck good

[ii] (e) [半 年 之内 船 不 坏 事]，
[bàn nián zhī nèi chuán bú huì shì]
[half year in boat NOT bad thing](embedding element),

[iii] (f) 于是 **他** 从 所赚 的 钱 上，
yūshì **tā** cóng suǒ zhuàn de qián shàng
so he from earned MOD money up,

(g) 又 讨 了 一 个 略 有产业 的
ø yòu tǎo le yí gè lüè yǒu chǎnyè de
(he) again marry PFV a M fairly well-to-do MOD

白脸 黑发 小寡妇。 （沈从文《边城》）
bái liǎn hēi fà xiǎo guǎfù
white-face black-haired young widow.

[Translation] He went home, though with bad feet. He bought a simple six-man wooden boat with his modest savings and rented it out to a boat captain … Luck was with him; the boat sailed safely, and in six months he'd saved money enough to marry a pretty, black-haired young widow. (Shen Congwen *Border Town*)

It is noted that several clauses led by the subject "he" are presented (omitted in the ellipsis [……] due to too many clauses in the original version). The embedment in part [ii], "bàn nián zhī nèi chuán bú huì shì", does not interfere with the control of its consequent comment clauses from the topic "tā (he)". Although inserted by other constructions, this topic chain is still very cohesive and unified.

Thirdly, as for some topic chains, a topic has two means of dominating each comment clause. For example,

(4) (a) **德国** 全境 受到 了 野蛮 战争 的 洗劫，
déguó quánjìng shòudào le yěmán zhànzhēng de xǐjié,
Germany whole land suffer to PFV savage fighting MOD loot,

(b) **市廛** 萧条，
ø **shìchán** xiāotiáo,
(Germany) town devastated,

(c) 田野 荒芜，
ø **tiányě** huāngwú,
(Germany) land desolated,

(d) 生灵 涂炭，
ø **shēnglíng** tútàn,
(Germany) people plunge into misery,

(e) 十 室 九 空。
ø **shí shì** jiǔ kōng
(Germany) ten houses nine empty

[Translation] In the savage fighting, Germany itself was laid waste, the towns and countryside were devastated and ravished, the people decimated.

Example (4) is conducted by the topic “Déguó” (Germany). Each clause in the chain is identified as a double-nominative construction when a zero form is recovered as the first NP. For example, the real form of “ø shìchán xiāotiáo” is “Déguó $_{NP1}$shìchán$_{NP2}$ xiāotiáo”. Within such a construction, a co-referential zero form usually occupies the initial position of each clause. Consequently, a zero form and the initial nominative entity form a double-nominative construction in each clause of the topic chain. As a common structure, double-nominative construction (DNC) often occurs in Chinese, which has been explored considerably (Shi, 2000; Chen, 2004; Sun, 2018). Its general pattern of a DNC is NP1, NP2 and a predicate following. Therefore, the topic “Germany” has a relation with “town”, “land”, “people” etc., in this way, when NP1 in each clause is omitted. The topic forms a whole-part, class-member, or possessor-possesee lexical semantic relation with NP2 in each clause and NP2 in each clause can be called a sub-topic.

Zero forms ahead of each sub-topic have a topic as their shared referential antecedent, which can well control each of its clause. Besides, a topic and its sub-topics have a meronymous[1] relation, both at the semantic-lexical level and the textual level, just as analysed in (4). The topic “Déguó” (Germany) has a meronymy relation with those items “shì chán, tián yě, shēng líng” in semantics, i.e. Germany encompasses town, land, people and ten houses, as its components. Therefore, the lexical-semantic connection and co-reference, conducted by the topic, together make this topic chain significantly unified and cohesive.

[1] We use “meronymy” to refer to part-whole relations. It is a kind of lexical relation studied carefully by Cruse (1986, pp. 157-180), and it was also analyzed as a lexical cohesion device in text firstly by Hasan (1984). Winston et al. (1987) propose six different types of meronymic relation: component-integral object, member-collection, portion-mass, stuff-object, feature-activity, and place-area. In Chinese discourse, meronymy is not only a kind of lexical cohesion device, but a device for connecting different clauses with the help of zero forms.

An English example is given to show a meronym in a discourse, "*The car* will not move. *The engine* is broken". Unlike English, a topic chain in Chinese uses different ways to combine its clauses, and as illustrated in Example (4), it prefers to use successive short parallel constructions (often four-character structure) containing subordinate words (town, land, people), which forms meronymous relationship with the superordinate word (Germany). Such a way of using a meronym in a topic chain can not be found in English or other languages (Lassalle & Denis, 2011).

3.1.2 The connection of comment clauses

Comment clauses within a topic chain are not arranged randomly; they are closely related in terms of one of the following three principles: cognitive orders, conjunction linking, and parallelism (or *pian-ou* construction).

Cognitive sequence, mainly embodied by the temporal consequence principle and spatial conceptual principle, is one general principle for Chinese to obey with respect to clause-clause and sentence-sentence combination. Tai (1985) explores the word order within a clause (sentence), proposing the Principle of temporal sequence (PTS), "The relative word order between syntactic units is determined by the temporal order of the states they present in the conceptual world". Tai's discussion is also applicable to the order of comment clauses within a topic chain. As shown in (5), comment clauses led by the topic "Xiao Zhang', a series of actions, are arranged in order of their occurrence in the physical world.

(5) (a) 小张 昨天 出 门 没 带 伞，
xiāo zhāng$_1$ zuótiān chū mén méi dài sǎn,
Xiao Zhang yesterday walk out NOT carry umbrella,

(b) 淋 到 雨 了，
ø$_1$ lín dào yǔ le,
(XiaoZhang) exposed to rain PFV,

(c) 受 了 风 寒，
ø$_1$ shòu le fēng hán,
(XiaoZhang) affect PFV wind chill,

(d) 发烧 躺 在 床 上 不 能 动 了，
ø$_1$ fāshāo tǎng zài chuáng shàng bù néng dòng le
(XiaoZhang) fever lie on bed up NOT can move PTCP

(e) 也 该 打个电话 请假 啊！
ø$_1$ yě gāi dǎ ge diànhuà qǐngjià a
(XiaoZhang) also should make a phone call ask for a leave MOD

[Translation] Without carrying any umbrella yesterday, Xiao Zhang was affected with chill after being exposed to the rain. Consequently, he could not move lying on the bed with a fever. In spite of this, he still should call to ask for a leave.

The comment clauses are stated according to the temporal order of event occurrence in Chinese. Their English translations, however, need not follow the original order due to many grammatical devices, such as conjunctions, relativization, etc., with which events are arranged flexibly in English.

Tai (1980) also mentions that, at the syntactic level, "Chinese tends to place the whole before the part, but English tends to do the reverse", and he holds "the whole-part relation is part of our (Chinese) perceptual system and is also a language universal principle". The general "whole-part" principle is quite easy to observe from Chinese word order, but how does the principle of whole-part relation effectively fulfil its function in the topic chain? Different from English, Chinese has few grammatical devices to implement the word order principle. As another cognitive sequence principle, the whole-part relation is supported by a lexical device to ensure that the type of a topic chain presented in (4) maintains orderly, meaningful and cohesive forms.

Developing an English paragraph by space is to arrange things according to their order of location and their relationship to each other. In a spatial sequence, information is arranged on the basis of geography or location, such as from east to west, from north to south, and so on. Take the description of digestion system, for example. In describing how we digest food, we begin with the mouth and work our way down the food pipe to the stomach and then to the intestine, and so on. The spatial sequence is commonly used to develop English paragraph, and it is also effective to arrange the order of comment clauses within the topic chain in Chinese. Temporal sequence and spatial sequence, as usual, are frequently used as cognitive order principles to organize Chinese sentences and discourse. Consider Example (6):

(6) (a) 中午 我 在 没有 导游 陪伴 的 时候
zhōngwǔ wǒ zài méi yǒu dǎo yóu péi bàn de shíhou
Noon, I at not have tour guide accompany de time

独自 漫步 街头，
dúzì mànbù jiētóu,
alone stroll street corner

(b) 在 中央 大道 附近 发现 了 一 个
ø zài zhōngyāng dàdào fùjìn fāxiàn le yí gè
(I) at Central Avenue nearby find PFV a M

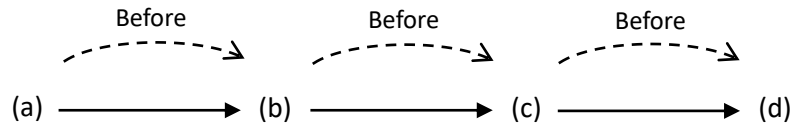
很 大 的 棚户区，
hěn dà de péngùqū,
very large de concentration of huts

(c) 很多 茅棚 里 还 住 了 人。
hěn duō máopéng lǐ hái zhù le rén.
many huts in still live PFV persons.

[Translation] Strolling unescorted at midday past a major concentration of the huts just a block from the city's Central Avenue I none the less saw many signs of occupation.

After "Central Avenue" precedes "a major concentration of huts", the locations described just follow the order of "observing", which is in accordance with the spatial order for Chinese readers/hearers, so all these geographical locations should be presented according to the logic order of "observing".

The relationship between comment clauses with zero forms is implicit. Consider the relationship among (a), (b), (c) and (d) in (5) by referring to its English translation. The clauses (a), (b) and (c) lead to the result "Xiaozhang could not move with a fever". The relationship among the first three clauses have been identified to be progressive cause-effect, that is, clause (1) directly causes clause (b), and clause (b) establish the cause of Xiaozhang's chill. Additionally, the four clauses are presented according to the temporal sequence. The following diagram shows implicit relationship among the four clauses in (5). No implicit conjunctions, connectives or other linking words can be found among the four comment clauses, but the topic chain can make the implicit relationship work out to tie with each clause closely.



Chinese tends to use quite a few conjunctions or connectives (or other linking devices) between clauses, in comparison with English which uses explicit conjunctions or connectives to join clauses. This tendency often occurs in a topic chain. The comment clauses within a topic chain are arrayed in their logical order in a physical world. Therefore, when we say that Chinese uses few conjunctions/connectives, the tendency is more appropriate for describing order of comment clauses within a topic chain. Sometimes semantic relationships between clauses are left implicit, so interpretations may sometimes require considerable creativity on Chinese speakers/readers. Take (5) as an example. There seems to be a missing link between clause (d) and clause (e), so Chinese natives probably predict that the missing link is about the fact that "Xiaozhang didn't go to work but failed to ask for a leave". The implicit semantic relationships between clauses often occur in a topic chain.

For complex actions and events, the cognitive order is sometimes difficult to perceive. In such cases, overt conjunctions or other linking words should be precisely inserted to indicate the complex relations. Normally, the reverse relation and other anti-cognitive relations should be specifically signified by the use of conjunctions,

shown as “què” (but) in (7). Observably, such conjunctions have a great influence on the topic expression forms (co-referential zero, pronominal) within a topic chain.

(7) (a) **五魁** 仰头 往 山 上 看，
wǔkuí$_1$ yǎngtóu wǎng shān shàng kàn,
Wukui raise head to hill top look,

看不到 岗梁
ø$_1$ kàn bú dào mǎo liáng
(Wukui) cannot see ridge

(b) **却** 想
ø$_1$ ***què*** xiǎng
(Wukui) but think,

(c) 若 立即 踏 桥 过 河， 山岗 上 必是
ruò Lìjí tà qiáo guò hé shān mǎo shàng bìshì
if Immediately step on bridge cross river, knoll up surely

能 看得见 的 了
néng kàn de jiàn de le
can be able to watch clearly DE PFV,

(d) 就 用 嘴 呶呶 左侧 的 一 处 鹰 嘴
ø$_1$ jiù yòng zuǐ náonáo zuǒcè de yí chù yīng zuǐ
(Wukui) MA Use mouth purse left side DE one M eagle beak

窝 岩，
wō yán
wall rock

(e) 说： [……] 要 站起来，
ø$_1$ shuō: yào zhàn qǐlái,
(Wukui) speak (Wukui) ready riseup,

(f) **却** 发现 自己 还 倒 在 草窝 里。
ø$_1$ ***què*** fāxiàn zìjǐ hái dǎo zài cǎowō lǐ
(Wukui) but find out himself still collapse into straw lair in

[Translation] Wukui lifts his head from the ground and looks up toward the top of the mountain. The ridge is obscured from his view. It occurs to him that if he should cross the bridge now, whoever is up there would be able to see him. He purses his lips and nods toward the left, indicating a cliff that juts out starkly from the wall of rock, like the beak of an eagle. He wants to rise, but finds himself collapsing into the straw lair.

Parallelism (or *pian-ou* construction)[2] is a frequently used method to organize Chinese clauses. “A parallel structure is a sequence of identical or near identical elements in the same corresponding positions in consecutive clauses” (Li, 2005, p. 113). Similarly, the device, parallelism or *pian-ou*, is also efficient to combine comment clauses, for example:

(8) 在 这里，**他** 比不上 橱窗 里 的 一 个
zài zhèlǐ, **tā**$_1$ bǐ bu shàng chúchuāng lǐ de yí gè
At here, he cannot match up shop window in DE one M

仿 古 花瓶，
fáng gǔ huāpíng
imitation ancient vase

比不上 掠身而过 的 一 身 紫色 的 衣裙，
ø$_1$ bǐ bu shàng lüè shēn ér guò de yì shēn zǐsè de yīqún
(he) cannot match up sweeping by DE one M purple DE skirt

更 比不上 牵在 女士 们 手 中 的
ø$_1$ gèng bǐ bu shàng qiān zài nǚshì men shǒu zhōng de
(he) still cannot match up pull in lady PL hand middle DE

那 条 小狗。
nà tiáo xiǎogǒu
that M dog

[Translation] Here, he is no match for an imitation of an ancient vase in the shop window, no match for the purple skirt sweeping by, and still no match for the little dog pulled by the hand of a lady.

The three comment clauses make use of the similar construction cannot match up + NP + Preposition + Modifier DE N to create repetition effect; accordingly, the three comment clauses are closely connected, largely by aid of the parallel structure.

Additionally, as four-character structure and parallelism are used properly, the Chinese topic chain will possibly become more readable because the two devices work effectively to strengthen the coherence among comment clauses. Look back at example (4). Comment clauses (b, c, d, and e) are presented by four-character

[2] The so-called symmetric sentence is two or two against the move, while the number of words is not required to be equal or symmetrical, and do not require similarly structured sentences. Symmetrical sentences were called *pian-ou* sentences in ancient China. Actually, *pian-ou*, similarly to parallel construction, functions as an effective syntactic mechanism to combine different syntactic units in Chinese. *pian-ou* is also a kind of hypotactic device to strengthen semantic cohesion in Chinese (Pan, 1997, pp. 352-355; Feng, 1997, pp. 133-144). Topic chains make use of the *pian-ou* device to strengthen combinations.

structure, such as “shìchán xiāotiáo”, “tiányě huāngwú”, etc., and they have formed a parallel relationship. As a whole, the passage in (4) therefore produces a coherent, powerful and musical effect with repetitious rhyme. However, the two strategies should be used in moderation, and too much abuse will result in undifferentiated effects, not only stale but wearisome.

3.1.3 The compactness of the topic chain

As mentioned at the beginning, readers of other languages will wonder why so much information is assembled into a short and compact structure and how the structure works effectively. We know that English will use several sentences and more linguistic devices to express the same amount of information that is stored in a Chinese topic chain with just several short clauses to form a chain-like structure. After so many clauses are compressed into such a compact structure like a narrow passageway, the subject (the topic) in each comment clause will be omitted and sometimes the same constituents adjacent to the subject (the topic) in the former clause will be invisible. Further, interestingly in many cases, Chinese has no need to use linking words in order to clarify temporal and logical relations between the clauses. These omissions may be caused by economic principles of the Chinese language. Due to such economic principles clauses are mostly arrayed according to natural and logical orders. When readers/hearers interpret a topic chain where much information is assembled into this narrow passage, they must heavily rely on their logical orders in physical world (such as PTS, spatial order, cause-effect, etc.) and the co-referential topic to find the implicit relationships, interpret their logical relationships and then understand the information in a topic chain.

In many cases the exact relationship between two clauses may be inferred from several optional logical orders which represent tacit knowledge for native Chinese speakers. However, sometimes the interpretation of implicit relationship needs creativity, inference or cognitive relevance. For example, when a clause (e), follows clause (d) in (5), the topic (Xiao Zhang) is still present in this clause, and therefore clause (e) is still one part of the topic chain though no linking words are used. Consider the relationship between (e) and other clauses. There is no implicit conjunction to indicate it, but readers/hearers are able to infer that (e) bears a concession relationship to the former four clauses. The concession relationship, different from the cause-effect relationship established by the former four comment clauses, is developed through the cognitive relevance to the former cause-effect and reasonable inference. If we use English to express the meaning, some linking word, such as “in spite of”, must manifest itself visibly.

As we know, when two English clauses are combined, an explicit linking word will be used to connect them, which makes the order of the two clauses relatively free. Meanwhile it is obligatory for English to indicate which clause is the main clause or

subordinate clause if these clauses are in subordination, that is, English requires one clause to establish an explicit grammatical and logical relationship with its adjacent one by the use of grammatical devices, such as conjunctions, non-finite verbs etc. By contrast, it is very hard to say which clause is the main clause in a topic chain in many cases, although these clauses bear some semantic relationship by means of their logical orders or reasonable references. Perhaps readers can judge which clause can be considered as the focus of information. Additionally, a Chinese topic chain seems to have an ability to continue only by adding short clauses if speakers want to extend the chain. For example, we can add more comment clauses followed after clause (e) in (5), "(f) ø (Xiao Zhang) zuì hào huán shì dào yīyuàn qù kàn kàn" (it's better for him to go to hospital to see the doctor.), "(g) ø (Xiao Zhang) nián jì bù xiào le, ø (Xiao Zhang) zěn me yě bù zhī dào zhàogù zì jǐa!" (He is not young, but why not know how to look after himself). If speakers still want to continue, more relevant comment clauses like (e), (g) can be added. These added clauses are still able to establish a close relationship with the former ones. In such a sense, this is one characteristic of the topic chain.

A topic chain can become a unified, meaningful and integrated construction with the help of a topic control and with close connection of clauses. Chinese tends to assemble amounts of information into very compact structure; in comparison with the Chinese topic chain, English is likely to use several sentences to express and indicate the small unit of subject-predicate by the use of finite verbs, as well as to demonstrate the referents of these zero forms. In such a sense, a topic chain is probably a quite convenient and effective tool to express large quantities of information through simple linguistic codes.

3.2 The external integration functions in discourse

Apart from having a powerful internal combination function, topic chains can effectively fulfil the function to construct discourse, through extending externally. In this section, the external integration functions of topic chains will be analysed from two aspects: how a topic chain links to other topic chains to develop discourse; how topic chains make use of non-chain constructions to advance discourse.

3.2.1 Alliance of topic chains

In Chinese, several different topic chains can be assembled together to construct discourse; resting on the association of several topic chains, a discourse comes into being. The alliance of topic chains can be viewed in two ways: as a sequence or as a hierarchy. When topic chains are organized as a sequence, this indicates that topic chains will occur in a sequence according to their chronological sequence or spatial order mentioned in 3.1. When topic chains are arranged as a hierarchy, it means that they are integrated into hierarchical whole-part relations. In accordance with their

linear characteristics and semantic relations, at least four kinds of modes can be summarized.

1) Mode of alternation between topic chains

A passage is filled with several topic chains, but only led by two topics. One topic and then the other alternate continuously, which often occurs in narrative passages. Consider the following example:

(9) [i] (a) 但 到 了 第二天， **人** 虽 起 了 床，
dàn dào le dì èr tiān, **rén**$_1$ suī qǐ le chuáng
but arrive PFV the second day, he although rise PFV bed

(b) ø$_1$ 头 还 沉沉 的。 **祖父** 当真
(Grandfather) tóu hái chénchén de. **zǔfù**$_1$ dàng zhēn
head still dazed PTCP Grandfather really

已病 了。
yǐ bìng le
get ill PFV.

[ii] (d) **翠翠** 显得 懂事 了 些，
cuìcuì$_2$ xiǎnde dǒngshì le xiē,
CuiCui seem sensible PTCP a little,

(e) ø$_2$ 为 祖父 煎 了 一 罐 大发药，
(CuiCui) wéi zǔfù jiān le yí guàn dà fāyào,
for grandpa concoct PFV a M medicinal herbs,

逼 着 祖父 喝，
ø$_2$ bī zhe zǔfù hē,
(CuiCui) force PTCP grandpa drink,

(f) ø$_2$ 又 在屋后 菜园地 里 摘取 蒜苗
yòu zài wūhòu càiyuándì lǐ zhāiqǔ suànmiáo
also behind the house vegetable plot in pluck garlic shoot

泡 在 米汤 里 作 酸蒜苗。
pào zài mǐtāng lǐ zuò suān suànmiáo
soak into rice soup in do sour garlic shoot

(g) ø$_2$ 一面 照料 船只，
yímiàn zhàoliào chuánzhǐ
at the same time take care boat

(h) ø$_2$ 一面 还 时时刻刻 抽空 赶回家
(CuiCui) yímiàn hái shíshí kèkè chōukòng gǎnhuíjiā
at the same time still hourly find time return home

里 来 看 祖父，
lǐ lái kàn zǔfù
in come see grandpa,

(i) 问 这样 那样。
ø$_2$ wèn zhèyàng nàyàng
(CuiCui) ask this that

[iii] (j) **祖父** 可 不 说 什么， 只是 为 一 个
zǔfù$_1$ kě bù shuō shénme, ø$_1$ zhǐshì wéi yí gè
Grandpa yet NOT say anything (Grandpa) only for a M

秘密 痛苦 着。 [……]
mìmì tòngkǔ zhe
secret suffer PTCP

[iv] (k) **翠翠** 看不出 祖父 有 什么 要紧 事情
cuìcuì$_2$ kàn bu chū zǔfù yǒu shénme yàojǐn shìqíng
CuiCui couldn't find out grandfather have what emergent thing

必须 当天 进城，
bìxū dàngtiān jìnchéng
must that very day go into town

(l) 请求 他 莫 去。 （《边城》）
ø$_2$ qǐngqiú tā mò qù
(CuiCui) request him NOT go

[Translation] Though he (grandfather) gets up next day, his head is still heavy. Cuicui, rising to the occasion, prepares a cooling concoction and makes him take it, after which she picks some garlic behind the house to boil with congee for him. Between trips on the boat, she runs home to see how he is. He says nothing, but his secret preys on his mind. Three days in bed restore him enough to walk about a little; and although his bones still ache, he decides to go into town. Cuicui can not understand what could be so important as to make Grandpa go to town so soon. She begs him not to go. (Shen Congwen, *biān chéng*)

The passage is composed of several topic chains ([i], [ii], [iii] and [iv]), but is led by just two topics "grandfather" and "CuiCui". The topic chains led by the two topics are switched easily and smoothly without the use of any grammatical marking or conjunction to strengthen the cohesion. The topic chains are arranged in terms of Temporal Sequence, so the way of association is viewed as a *sequence* mentioned at the beginning of this part; as a result, the discourse woven by two kinds of topic chains is quite natural and fluent. Based on the mode (two or three topics alternate repeatedly), several topic chains are successively connected to construct narrative discourse, which especially occurs in novels.

2) Mode of meronymous topic chains

After one topic chain comes to an end, other topics in the successive topic chains have some semantic relations with the topic of the previous topic chain. Semantic relations are expectedly meronymous, expressing either whole-part, class-member, or possessor-possesee relation. For example:

(10) [i] (a) **泸沽湖** 是 一 个 天然 内陆 淡水湖，
lúgū hú₁ shì yí gè tiānrán nèilù dànshuǐ hú
Lugu Lake is a M natural inland freshwater lake

位于滇 西北 云南 与 四川 两 省 交界处，
wèiyūdiān xīběi yúnnán yǔ sìchuān liǎng shěng jiāojièchù
locate in NW Yuannan and Sichuan two provinces juncture

[ii] (b) (湖) 面积 50 平方公里，
[ø₁ miànjī 50 píngfāng gōnglǐ,
(Lugu lake) area 50 square kms

(c) 海拔 2680 米，
ø₁ hǎibá 2680 mǐ
(Lugu lake) sea level 2680 meters

(d) (湖) 平均 水 深 45 米。
ø₁ píngjūn shuǐ shēn 45 mǐ]
(Lugu lake) average water depth 45 meters,

[iii] (e) **湖中** 有 八 岛 十四 湾 和 一 个 海堤连岛，
húzhōng₂ yǒu bā dǎo shísì wān hé yí gè hǎidī liándǎo
lake middle have 8 islands 14 coves and a M sea bank dyke

(f) 小岛 棋布星罗；
ø₂ xiǎo dǎo qíbù xīngluó
(lake middle) small island spread all over the place

[iv] (g) **湖岸** 植被 葱郁，
hú'àn₃ zhíbèi cōngyù,
lake bank plants verdant,

(h) 青山 环绕， 风光 旖旎。
ø₃ qīngshān huánrào ø₃ fēngguāng yǐnǐ
(lake bank) green hills surround (lake bank) scenery exquisite

[Translation] *Lugu* Lake is a natural freshwater lake located at the juncture of southwest China's Yunnan and Sichuan Provinces, with an altitude at 2680 meters above sea level and an average depth of 45 meters. It covers an area of 50 square km in which there are eight islands, 14 coves and a dyke, and small islands

spreading throughout the lake, with verdant plants thriving on its banks. Surrounded by green hills, the lake shows exquisite scenery.

Part [ii] in Example (10) is assembled by several double nominal constructions, having several co-referential zero forms with the topic "*Lugu* Lake". Since part [i] is defined as a zero-form topic chain, [i] and [ii] together make a larger topic chain. The same pattern recurs in two other topic chains [iii] and [iv] which have different topics, led by 'hú zhōng' (lake middle) and 'hú àn' (lake bank) respectively. Obviously, entities of the two topics and co-referential zero forms in [iii] and [iv] can be regarded as components of "*Lugu* Lake", which makes us conclude that the first topic, from the lexical relation perspective, has set up a merynomous relation with the latter two topics. Topic chains are thus organized hierarchically, and compared to the first topic chain conducted by "*Lugu* Lake", the second (lake middle) and third topic chains (lake bank) appear at lower ranks. .

In this way, topics establish a meronymous relation lexically, with the three different topic chains being integrated into a hierarchical whole-part relation. The example illustrates that different topic chains with meronymy relations can build discourse efficiently.

3) Mode of network form

Generally, linear characteristics of topic and comment clauses can be described as the topic being the focus, and comment clauses as its radiated items, A traditional zero-form topic chain therefore can be drawn into a radioactive diagram. When several zero form topic chains occur successively, they are woven into a network-shape construction as shown in (11):

(11) [i] (a) 小王 放下 书包，
Xiǎo Wáng$_1$ fàngxià shūbāo
Xiao Wang put down bag,

(b) 把 [ii] **球** 扔过去，
ø$_1$ bǎ **qiú**$_2$ rēng guòqù
(Xiao Wang) BA ball throw pass over

(c) 正好 砸到 窗户，
ø$_2$ zhènghǎo zádào chuānghu
(ball) just fall to window

(d) （球把） 玻璃 打坏 了，
ø$_3$ bōli dǎhuài le
(ball make) glass broken PFV,

[iii] (e) 碎片 伤 了 **小明**，
suìpiàn shāng le **Xiǎo Míng**$_4$
fragments hurt PFV Xiao Ming

(f) **手** 破 了，
ø$_4$ **shǒu**$_5$ pò le
(Xiao Ming) hand injure PFV

(g) （小明手） 流 了 [iv] 不少 **血**，
ø$_5$ liú le bù shǎo **xuè**$_6$
(Xiao Ming hand) lose PFV a lot of blood

(h) 淋 到 衣服 上，
ø$_6$ lín dào yīfu shàng
(blood) flow to clothes up

(i) （血把衣服印） 红 了 一 大 片，
ø$_6$ hóng le yí dà piàn
(blood) dye red PFV a large piece

(j) 怎么 也 止不住，
ø$_6$ zěnme yě zhǐbuzhù
(blood) anyway also cannot stop

[v] (k) （小明） 哇哇 大 哭起来。
ø$_4$ wāwā dà kū qǐlái
(Xiao Ming) onomatopoeia big cry up

[Translation] After putting down his bag, Xiao Wang threw the ball away, but the ball just hit the window; as a result, the window glass broke into pieces. The broken glass fragments injured Xiao Ming' hands and he lost quite a lot blood. The blood spread into his clothes and largely dyed them red. Anyway, the blood could not be stopped, so Xiao Ming cried loudly.

The example above displays a network formed by several different topic chains. The passage consists of several zero form topic chains, and each topic chain is advanced progressively ([i], [ii], [iii], [iv], [v]). The eight topics are mentioned successively in the passage: *Xiao Wang—qiu—chuan wai—sui pian—Xiao Ming—shou—xue—Xiao Ming*. The second topic ([ii]) is mentioned in the comment clause in the first topic chain([i]). After the first topic chain terminates, the second topic begins to form its chain. A similar pattern recurs with all further topic chains, and we can say that the previous chain is transited onto the next, and therefore all topic chains together form a network.

Through transiting from one to the next, each topic is related with the next consistently and smoothly. Finally, these topic chains are woven into a complex network, shown in Figure 1:

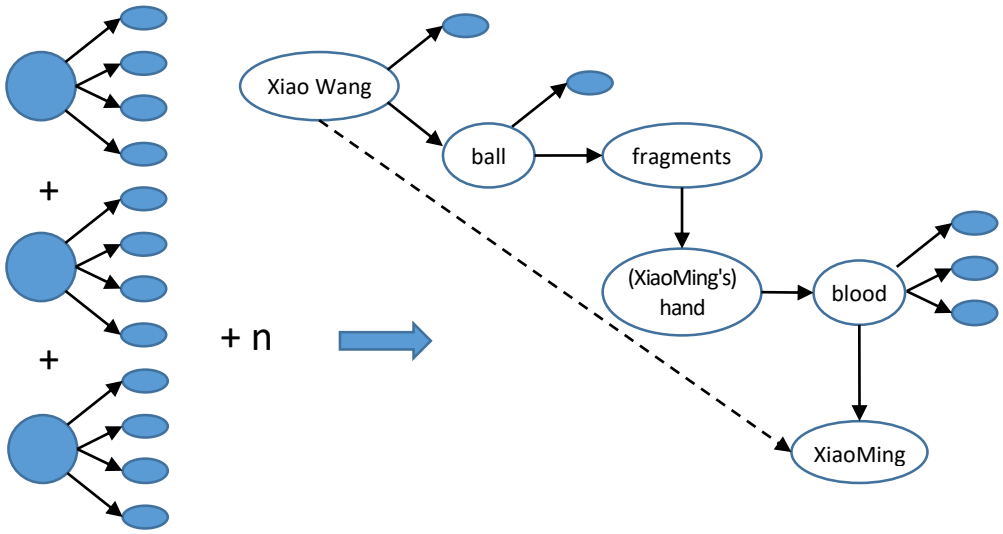


Figure 1: Representation of combination process for Example (11)

In this way one topic chain is associated with the sequential topic chain through the transitional progression; consequently, different topic chains are interwoven into a larger network, but the general order of arrangement for the sequence of topic chains should be in accordance with PTS. The two means ensure that cohesion in the passage is substantially enhanced.

Noticeably, Chu (1998, pp. 330-337) mentions two types of a topic chain: the embedded topic chain and the telescopic topic chain. The embedded topic chain is the performance of one topic chain permitting other structures to enter, which also proves strong domination of the topic over the whole chain. A telescopic topic chain is defined as involving two topic chains merging into each other at the end of one and at the beginning of another chain. Just as a discourse pivot serves as the object of the first verb and the subject of the second verb, it also serves as the last link of one chain and the first link of another. The piece of discourse that consists of two or more topic chains linked in this manner can be called a telescopic topic chain. From Chu's examples, we find that a telescopic topic chain coincides with the network formed by different topic chains through the transited manner. The transited connected manner is more complex than the pivot which serves as the object of the first verb and the subject of the second verb, so a telescopic topic chain is a special case of our network mode. In this way, the mode in the study is more inclusive than the telescopic topic chain.

Additionally, Cheng (1988) proposes a model of topic continuity which You (1998, p. 32) bases on in the following diagram. The model is similar to the mode in this study, showing that this mode of network form is a continuation of Cheng's and Chu's work.

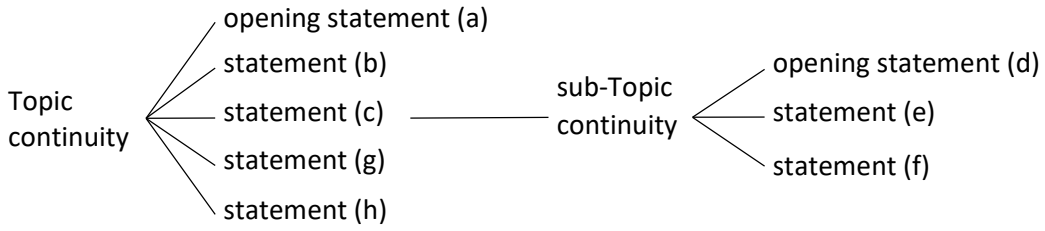


Figure 2: The model of topic continuity (Cheng 1988)

4) Embedding modes of topic chains

Embedded topic chain (Chu, 1998, p. 330) can be considered as the performance of a topic chain permitting other structures to enter, where the embedded topic chain is regarded as a part of a larger chain. At the same time, an embedded topic chain and its larger topic chain can also be viewed as the two separate topic chains, so the embedded relation can also be defined as a mode of different topic chains establishing discourse. For example,

(12) (**李四** 这 家伙，**我** 因为 救 他 受 了 伤，
(**Lǐsì**$_1$ zhè jiāhuo [**wǒ**$_2$ yīnwèi jiù tā ø$_2$ shòu le shāng,]$_b$
Lisi this dude I because save him, (I) receive PFV wound,

竟然 不 来 看 我，
ø$_1$ jìngrán bù lái kàn wǒ
(Lisi) even NOT come see me

跑到 纽约 度假 去 了。
ø$_1$ pǎodào Niǔyuē dùjiǎ qù le.)$_a$
(Lisi) run-to New York have vacation go PFV

[Translation] Lisi, that dude I was wounded in saving, doesn't even come to see me and went to New York for a vacation.

The topic chain controlled by "wǒ" (I) is embedded within a larger topic chain conducted by "Lisi (a name of a person)". As stated in 3.1, a topic chain is tolerant to being inserted by other syntactic constructions, and the inserted syntactic construction is possibly identified as another topic chain. So the language phenomenon can be described as a large topic chain having a small, different one attached. This is also one method for topic chains to create textual coherence.

Concerning this mode, we should pay special attention to the change in antecedents of zero forms. In Example (12), the first zero form refers to "I", but the second zero has the antecedent of "Lisi", the same as the third zero. The question that arises is how speakers/readers know that the referents of the second and third zero form have been shifted. It is the clause, "ø$_1$ jìngrán bù lái kàn wǒ", that provides many

cues for speakers/readers to know who performs the actions. The predicate “not come to see me”, for example, denies that “I” will come to see me, so only “Lisi” will be qualified to take the action. Since “I” was wounded, it is impossible for “me” to go to New York, and thus the third zero must refer to “Lisi”. The continuous zero forms having different referring entities can be judged from the clues provided from the context, and are called switch referents (You, 1998; Lee, 2003; Tao, 1996, 2001).

Primarily dependent on the four modes, topic chains take full advantage of lexical semantic linkage and cognitive consequence principles to constitute Chinese discourse. Compared with other languages, it is distinctive of Chinese to construct discourse through the four modes by using few grammatical devices. This has not been found in English (Esser, 2009) or other European languages (Longacre, 2007, pp. 372-420) so far.

3.2.2 The incorporation of topic chains and non-chain forms

It is almost impossible for Chinese discourse to be entirely composed of one topic chain alone. Once one topic chain determinates, it is likely to be followed by other topic chains, and sometimes by other constructions. In the realistic context, one topic chain is possibly subsequently connected with other non-chain constructions. Observe the following examples.

(13) (a) 船夫 方面 还 以为 上次 歌声 既归
(T1) **chuánfū**$_1$ (C1) fāngmiàn hái yǐwéi shàngcì gēshēng jì guī
Ferryman (T1) side (C1) still think last time song attribute

二老 唱 的，
èrlǎo chàng de
the second sing DE

(b) 在此后几个日子里， 自然 还 会 听到 那 种 歌声。
zài cǐ hòu jǐ gè rìzi lǐ zìrán hái huì tīngdào nà zhǒng gēshēng
In the next few days in, should still can listen to that kind song

(c) 一到 了 晚间 就 故意 从 别样 事情 上，
yí dào le wǎnjiān jiù (C2) gùyì cóng biéyàng shìqíng shàng
arrive PTCP evening MA intentionally from other things up,

(d) 促 翠翠 注意 夜晚 的 歌声。
ø$_1$ (C3) cù cuìcuì zhùyì yèwǎn de gēshēng
(ferryman)(C3) urge CuiCui notice evening DE song.

(e) [两 人 吃 完 饭 坐 在屋里，
[liǎng rén chī wán fàn zuò zài wū lǐ
[Two persons eat PFV dinner sit in the house,

	因	屋前	滨水，
	yīn	wū qián	bīn shuǐ
	because	front near	water,

(f)	长脚	蚊子	一到	黄昏	就	嗡嗡	的	叫着]，
	chángjiǎo	wénzi	yídào	huánghūn	jiù	wēngwēng	de	jiào zhe]
	long leg	mosque	to arrive	musk	MA	buzz	DE	drone]

(g)	**翠翠**	便	把	蒿艾	束成	的	烟包	点燃，
	cuìcuì$_2$ (T2)	biàn	bǎ	hāoài	shùchéng	de	yānbāo	diǎnrán
	CuiCui(T2)	then	BA	Artemisia	bunch into	DE	smoke coil	light

(h)		向	屋	中	角隅	各处	晃	着
	ø$_2$ (C1)	xiàng	wū	zhōng	jiǎoyú	gèchù	huǎng	zhe
	(CuiCui) (C1)	to	house	middle	corner	every place	wave	PTCP

	驱逐	蚊子。
	qūzhú	wénzi
	drive	mosquito.

(i)		晃	了	一阵。	（《边城》）
	ø$_2$ (C2)	huǎng	le	yízhèn	
	(CuiCui)(C2)	wave	PFV	a while	

[Translation] The ferryman thinks that, since Number Two was the singer, they will hear more of his songs in the next few days. When evening falls, he encourages Emerald to listen for songs that night. After supper they sit indoors. Because their hut stands above a stream, it is filled with the droning of mosquitoes at dusk, and Emerald waves a lighted coil of Artemisia in every corner to drive the mosquitoes away. (Shen Congwen *biān chéng*)

There are two types of topic chains constructed by “T1 (ferryman)” and “T2 (CuiCui)” respectively in the passage. Nevertheless, other non-chain forms marked in grey in the passage are embedded within the two topic chains, and the embedment (clause e and f) performs the background[3] to transit from the first chain to the second. Although the non-chain constructions are not directly related to the time of the narrative, which has a clear line of progression formed by topic chains in the passage, the constructions, as the background information, still link the two events stated in the two topic chains.

[3] Thompson (1987) and Givón (1987) published a detailed analysis of a background and foreground, but so far the relations of the two terms have not been clearly stated. They state simply, a background is often considered to be the material that represents sidetracks and thus does not have to be in temporal order, generally takes stative verbs, and is usually coded in imperfective aspect. By contrast, a foreground is regarded to be the material that presents the event line of a narrative. (Chu, 1998, p. 219).

(14) (a) 五魁 便 见 天清早 拾 粪， [……]，
wǔ kuí biàn jiàn tiān qīng zǎo shí fèn
Wukui MA see morning gather animal droppings

(b) 看 着 柳 家 门 前 的 动静，
ø kàn zhe liǔ jiā mén qián de dòng jing
(Wukui) look PTCP Liu house gate front DE come and go.

(c) 终一日， 太阳 还 没有 出来，
[zhōng yírì tàiyáng hái méiyǒu chūlái
One day, sun still NOT come out

(d) 村口、 河岸 一层 薄 雾 闪动 着 蓝光。
cūn kǒu, hé àn yì céng báo wù shǎndòng zhe lánguāng.]
village entrance, river bank a layer thin mist waver PTCP blue light

(e) 五魁 瞧见 女人 提 着 篮子 到 河边
[wǔ kuí qiáojiàn nǚrén tí zhe lánzi dào hébiān
Wukui see woman carry PTCP basket arrive riverbank

洗 衣服 了。
xǐ yīfu le]
wash clothes PFV.

(f) 女人 还是 那么 俊俏， (她的) 脸 却 苍白 了 许多，
nǚrén háishì nàme jùnqiào liǎn què cāngbái le xǔduō
woman still so pretty, (woman) face but pale PFV very much

(g) 挽 了 袖子 将 白藕 般 的 胳膊 伸 进
ø wǎn le xiùzi jiāng bái ǒu bān de gēbo shēn jìn
(she) roll PFV sleeve JIANG white lotus root like DE arm put into

水 里 来回 搓摆。 （《五魁》）
shuǐ lǐ láihuí cuōbǎi
water in repeatedly twist

[Translation] He takes up the practice of gathering animal droppings at the crack of dawn…. Or he stands afar on the opposite bank of the river that flows past it, watching the comings and goings. Finally, one day, before the sun has arisen, when the entryway to the village and the river as well are both bathed in a luminous bluish mist, he sees the woman carrying a basket of laundry to the riverbank. She's as pretty and charming as ever, though her face is paler than he remembers. After rolling up her white lotus root-like arms, she puts her arms into water to twist away clothes repeatedly. (*wǔ kuí*)

The passage consists of two topic chains, (a)–(b) and (f)–(g), and two non-chain constructions, (c)–(d) and (e). The non-chain construction (c)–(d) does not participate

in the narrative process of the passage as background information, while the non-chain structure (e) manages to connect one topic chain (a)–(b) and the other chain (f)–(g) together through its participation into their shared narrative progression and event-line time. At the moment, non-chain construction (e) works as the foreground information to advance the narrative progression, and strengthen the combinations of two topic chains as well as narrative and non-narrative information.

The functions of non-chain constructions can be summarized as:

1) the non-chain construction is encoded as background information, which is closely relevant with what is discussed in the topic chain, but a non-chain construction does not participate in the narrative process of topic chains, shown as in (13);

2) the non-chain construction also serves to associate one topic chain with the other through its narrative participation and time trace, and sometimes the construction possibly participates in the time process of topic chains, as in (14).

4 Conclusion

Clearly, the topic chain in Chinese can construct discourse and help achieve textual coherence. The topic chain in Chinese discourse has its own special mechanism for combining different discourse units into a large textual segments, which is only found in a few languages, particularly in East Asian languages.

The study concentrated on the integration functions of the topic chain in Chinese discourse. To clarify how topic chains integrate smaller units into larger ones in discourse, many efforts have been made to investigate its internal unification and external integration functions in discourse. The study tentatively presented several discourse organizational modes of topic chains, illustrated by many realistic Chinese texts. It carefully and closely examined the integration functions for the topic chain, and is thought to be significant for deeper understanding the nature of topic chain and how discourse coherence is established in Chinese.

References

Chen, P. (2004). Hanyu shuangxiang mingciju yu huati chenshujiegou 汉语双向名词与话题-陈述结构 [Double NP constructions and topic-comment articulation in Chinese]. *Zhongguo Yuwen 中国语文 [Studies of the Chinese Language*], 6, 493-507.

Cheng, C-C. (1988). Tongxin benwei hanyu pianzhang yufa 通信本位汉语篇章语法 [Communication-based Chinese discourse grammar]. *Shijie Hanyu Jiaoxue 世界汉语教学 [Chinese Teaching in the World*], 1, 6-13.

Chu, C-C. (1998). *A Discourse Grammar of Mandarin Chinese*. New York: Peter Lang.

Cruse, A. D. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

Dixon, R. M. W. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Esser, J. (2009). *Introduction to English Text-linguistics*. Frankfurt: Peter Lang.

Feng, S. (1997). *Hanyu de Yunlü, Cifa yu Jufa 汉语的韵律、词法与句法 [Interactions between Morphology, Syntax and Prosody in Chinese]*. Beijing: Peking University Press.

Givón, T. (1987). Beyond foreground and background. In R. Tomlin (Ed.), *Coherence and Grounding in Discourse* (pp. 173-188). Amsterdam: John Benjamins.

Halliday, M.A.K & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hasan, R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.) *Understanding Reading Comprehension* (pp. 181-219). Delaware: International Reading Press.

Lassalle, E. & Denis, P. (2011). Leveraging different meronym discovery methods for bridging resolution in French. In I. Hendrickx, S., Devi, A., Branco & R., Mitkov (Eds.) *Anaphora Processing and Applications (pp.* 35-46). Berlin: Springer.

Lee, C-L. (2003). *Zero Anaphora in Chinese*. Taipei: Crane Publisher.

Li, C. N., & Thompson, S. A. (1976). Subject and topic: a new typology of language. In C. N, Li. (Ed.) *Subject and Topic (pp.* 457-61). New York: Academic Press.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Li, I-C. (1985). *Participant Anaphora in Mandarin Chinese*. PhD dissertation, University of Florida.

Li, H. (1995). *Topic Chain Structure in Chinese Conversations*. PhD Dissertation, University of Minnesota.

Li, W. (2004). Topic chains in Chinese discourse. *Discourse Processes,* 37(1), 25-44.

LI, W. (2005). Topic Chains in Chinese: A Discourse Analysis and Applications in Language Teaching. Muenchen: Lincom Europa.

Longacre, R. E. (2007). Sentences as combinations of clauses. In T. Shopen (Ed.) *Language Typology and Syntactic Description* (2nd edition) Vol.2: Complex Structure (pp. 372-420). Cambridge: Cambridge University Press.

Pan, W. (2007). *Hanyingyu Duibi Gangyao 汉英语对比纲要 [The Outline of Contrastive Studies between Chinese and English]*. Beijing: Beijing Language and Culture University Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Shi, D. (1989). Topic chain as a syntactic category in Chinese. *Journal of Chinese Linguistics*, 17, 223-262.

Shi, D. (1992). *The Nature of Topic Comment Constructions and Topic Chains*. PhD dissertation, Southern California University.

Sun, K. (2018). Approaching the double-nominal construction in Mandarin Chinese through the semantic-cognitive interaction. *Studia Linguistica*, *72*(3), 687-724.

Tai, H.Y. (1980). Toward a cognition-based functional grammar of Chinese. In H. Tai & H. Frank (Eds.), *Functionalism and Chinese Grammar no.1* (pp.187-226), Monograph Series of the Journal of the Chinese Language Teachers Association.

Thompson, A. S. (1987). "Subordination" and narrative event structure. In R. Tomlin (Ed.) *Coherence and Grounding in Discourse* (pp. 435-454). Amsterdam: John Benjamins.

Tao, L. (1996). Topic discontinuity and zero anaphora in Chinese discourse: cognitive strategies in discourse processing. In B. Fox (Ed.), *Studies in Anaphora* (pp. 487-513). Amsterdam: John Benjamins.

Tao, L. (2001). Switch reference and zero anaphora: emergent reference in discourse processing. In A. Cienki, B. Luka & M. Smith (Eds.), *Conceptual and Discourse Factors in Linguistic Structure* (pp. 253-269). Stanford: CSLI Publications.

Tsao, F-F. (1979). A Functional Study of Topic in Chinese: The First Step Towards Discourse Analysis. Taipei: Student Book Co.

Tsao, F-F. (1990). Sentence and Clause Structure in Chinese: a Functional Perspective. Taipei: Student Book Co.

Winston, M.E., Chaffin. R. & Herrman, D. 1987. A taxonomy of part-whole relation. *Cognitive Science*, 11(4), 417-444.

You, Y-L. 1998. *Interpreting Chinese Zero Anaphora Within Topic Continuity*. PhD dissertation, University of Illinois at Urbana-Champaign.

Abbreviations and symbols

CONJ	conjunction
DE	modifier (de)
MSTC	meronymy style topic chain
M	measure word
MA	modality adverbs (jiu, cai, you, hai, etc.)
MOD	modifier
Ø	zero form
PASS	passive
PROG	progressive
PFV	perfective aspect
PTCP	particle
ZA	zero anaphor

Acknowledgments

This study was funded by the National Social Sciences Foundation of China (Fund No. 15CYY038) and China Postdoctoral Science Foundation (Fund No. 2018T110581).

TRACING THE IDENTITY AND ASCERTAINING THE NATURE OF BRAHMI-DERIVED DEVANAGARI SCRIPT

Krishna Kumar PANDEY
Indian Institute of Technology Roorkee, India
Department of Humanities and Social Sciences
krishnapandeybuxar@gmail.com

Smita JHA
Indian Institute of Technology Roorkee, India
Department of Humanities and Social Sciences
smitaiitr@gmail.com

Abstract

Current research exploits the orthographic design of Brahmi-derived scripts (also called Indic scripts), particularly the Devanagari script. Earlier works on orthographic nature of Brahmi-derived scripts fail to create a consensus among epigraphists, historians or linguists, and thus have been identified by various names, like semi-syllabic, subsyllabic, semi-alphabetic, alphasyllabary or abugida. On the contrary, this paper argues that Brahmi-derived scripts should not be categorized as scripts with overlapping features of alphabetic and syllabic properties as these scripts are neither alphabetic nor syllabic. Historical evolution and linguistic properties of Indic scripts, particularly Devanagari, ascertain the need for a new categorization of its own and, thus preferably merit a unique descriptor. This paper investigates orthographic characteristics of the Brahmi-derived Devanagari script, current trends in research pertaining to the Devanagari script along with other Indic scripts and the implications of these findings for literacy development in Indic writing systems.

Keywords: orthography; Brahmi; Devanagari, akshara; alphasyllabary; alphabet

Povzetek

Raziskava obravnava ortografsko obliko pisav, ki izhajajo iz pisave brahmi (imenovane tudi *indijske pisave*), še posebej pisavo devanagari. Predhodnje študije o ortografski naravi the pisav niso uspele povezati mnenj epigrafov, zgodovinarjev in jezikoslovcev, zato je njihov opis zelo raznolik; uporabljajo se poimenovanja kot npr. polzlogovna, podzlogovna, alfa-zlogovna oz. *abugida*. V nasprotju s tem članek zagovarja idejo, da

Acta Linguistica Asiatica, 9(1), 2019.
ISSN: 2232-3317, http://revije.ff.uni-lj.si/ala/
DOI: 10.4312/ala.9.1.59-73

pisav, ki izvirajo iz pisave brahmi ne bi smeli označiti s podvajajočimi se značilnostmi abecenih in zlogovnih pisav, saj niso ne abecedni in ne zlogovni. Zgodovinski razvoj in jezikoslovne lastnosti indijskih pisav, še posebej pisave devanagari, nakazujejo na potrebo po oblikovanju nove kategorije, ki bi jo lahko poimenovali 'aksarske pisave'. Za konec članek ponuja kratek pregled razvoja pismenosti na področju aksarskih pisav.

Ključne besede: pisava, pisava brahmi; pisava devanagari, aksara; alfa-zlogovnica; latinica

1 Introduction

It has long been argued that 'pictures' can be quoted as the first instance of a kind of writing system. Interestingly, Gelb (1963), in his monumental work, categorises pictures under the first stage of writing, called "No Writing." He claims that a picture, which is an object of art, results from an artistic-aesthetic urge that fails to support theories of writing systems. However, under the heading of "Forerunners of Writing", he coined a term 'semasiography' which shows the stage in which pictures (here, he differentiated between artistic pictures and simple pictures) can convey general meanings. Certainly, Brahmi, an ancient Indic script does not make its appearance either in the category of 'no writing' or in 'semasiography'. It comes under the phonography, a category representing fully developed writing systems. Nevertheless, a question mark has always been put on the nature or identity of Brahmi or Brahmi-derived scripts. Gelb (1963, p. 187) writes that "the forms of the individual signs of the Brahmi writings show no clear relationship with any other system, and were most probably freely invented." With these words, he raises a fundamental question on the nature of the linguistic organisation of Brahmi. The aim of this paper is to investigate orthographic characteristics of the Brahmi-derived Devanagari script, current trends in research pertaining to Brahmi-derived scripts and the implications of these findings for literacy development in akshara based writing systems.

2 Theoretical Background

It is well established that phonological structure plays a major role in defining the writing system of a language. The stream of sound segments of a spoken language is not perceived discretely but can be artificially segmented into individual phonological units. Syllable, a cluster of sounds, is a hierarchically structured phonological unit, which comprises an onset, a nucleus, and a coda to constitute different sound sequences of a language. In a syllable, nucleus is an obligatory component, while onset and coda are optional components. Of these, syllables without a coda are open

syllables and syllables with a coda are closed syllables (Castles & Coltheart, 2004; Gordon & Ladefoged, 2001; P. Pandey, 2007).

Further, within a phonological system consonants and vowels act differently. Thus, the root form of a word is composed of two or three consonants, while vowels impart different grammatical aspects to it. This phenomenon is easily perceptible in Semitic languages (McCarthy, 1981). The theory of Dependency Phonology proclaims that human speech comprises three basic vowels – /a/, /i/, and /u/, and others are produced from their amalgamation (Anderson & Ewen, 1987). In addition, Government Phonology states that consonants carry an inherent short vowel, which is suppressed by individual languages in which it does not surface; otherwise, it surfaces as /ə/ (Kaye, Lowenstamm, & Vergnaud, 1985).

Orthography signifies writing system of a language. Structures of different orthographies vary at the level of phonological awareness they represent and thus it can be assumed that orthographic domain is shaped by the nature of its writing system (Ziegler & Goswami, 2005). Orthographies of various languages are controlled by different factors, e.g. Hindi orthography, Arabic orthography, and English orthography depend on phonological awareness, lexical awareness, and morphological awareness respectively (Pandey, 2007). With categorization of the nature of orthographies, syllabaries are such phonetic writing systems that represent the phonological units at the level of syllables. The Japanese Hiragana, for example, represents the syllable sound sequence /ka/ with the symbol か, /ki/ as き and /ku/ as く. Characters for /ka/, /ki/, and /ku/ in Japanese hiragana have no similarity to specify their common sound /k/. On the other hand, alphabetic writing systems represent sounds at the smallest pronounceable segment of speech which is a phoneme. Thus, a syllable /ka/ represents two graphemes of English, i.e., 'k' and 'a'. Indic writing systems, on the contrary, represent phonological units at both the syllabic and alphabetic levels concurrently (Bright, 1996; Nag, 2011).

3 History and description of Brahmi

In India, Brahmi evolved and flourished around third century B.C.E. during the Ashokan regime (272-326 B.C.E). The edicts of Ashokan period extensively represents the Northern-Brahmi script (Verma, 1971). Experts of Paleography have primarily considered the Ashokan Brahmi a fully matured writing system. As Upasak (1960, p. 21) explains,

> "Brahmi may have begun as a mercantile alphabet, based either on vague memories of the Harappa script or derived from contact with Semitic traders, indeed it may have owed to both these sources; but by the time of Ashoka, it was the most developed and scientific script of the world."

Similarly, Basham (1967) argues that the documentation of Brahmi script, to represent the Sanskrit phonology in the Ashokan inscriptions, shows a rich and long developmental history of it.

Brahmi has been linguistically adapted for the genesis of several scripts in the area of Indian subcontinent as well as in South-East Asia. The scripts used for writing Indo-Aryan, Dravidian, Austro-Asiatic, Tibeto-Burman trace their roots back to the Brahmi. Hindi, Gujarati, Tamil, Telugu, Kannada, Malayalam, Bengali, Assamese, Punjabi of India, Sinhala in Sri Lanka, Tibetan, Javanese, Khmer, Thai, Burmese in South East Asia use scripts based on Brahmi (Gelb, 1963; Ruhlen, 1991). Roop (1972, p. ix) states that "the extent of early Indian influence in continental South-East Asia is nowhere more apparent than in the use of Indian writing systems for noncognate languages covering large parts of the latter area."

The outset of the architecture of Brahmi can best be perceived by making an outright connection with the linguistic design of the oral mode of learning in Vedic India. Knowledge, in the Vedic time (5000 B.C.E.), was transferred from one generation to the next through the tradition of *Shruti* (hearing) and *Smriti* (memory). Shruti, referring four Vedas, is created in the language of *Vedic* Sanskrit having a fixed accent which was used to converse in musical notes. This oral tradition authorises nobody to make a single change of the Shruti even at the level of a syllable. This rigidity made disciples learn the Shruti (hence, the literature hearing or Vedas hearing) with acute phonetic precision. While, Smriti, written in *laukika* Sanskrit, has been defined as the literature composed by self-realization of sages whose fundamental thoughts are primarily based on the comprehension of the Shruti (Kapoor, 2002; Mukherjee, Nema, & Venkatesh, 2012). Scharfe (1977, p. 130) writes, "The Veda reciter had to learn how to constitute the continuous text from the word-for-word text, observing the rules of vowel and consonant sandhi as well as those of accentuation." Based on these facts we can argue that the Shruti and Smriti tradition have hugely affected the structure and design of Brahmi and the scripts derived thereafter.

Theories propounded to trace the origin of the Brahmi script have broadly been divided into two groups: 1) theories that associate their origin with an indigenous source, and 2) theories that trace their origin from some foreign source. The theory of the indigenous origin of Brahmi includes scholars like, Lassen and Edward Thomas, who credited the origin of Brahmi to the Dravidian races of South-India (Cited in Upasak, 1960). This assumption was probably based on the Aryan-Dravidian theory. Historians claim that Dravidians inhabited entire India before the advent of Aryans in this land. Also, Dravidians were culturally more advanced than Aryans, hence invented the writing system much before the Aryan's settlement (Pandey, 1957). Since the theory was based on presumptions, it fails to get the proper recognition from the esteemed scholars. Among others, Pandit G.H. Ojha (1959, cited in Upashak 1960, p. 13) very strongly asserts in his books that "Brahmi letters were developed in India out of

pictographs and were later perfected to best suit the phonological character of the languages. No foreign influence can possibly be traced through the formation of letters." Another supporter of the indigenous origin of Brahmi, an Indian scholar R. Pandey (1957, p. 50) advocates that "… Brahmi characters were invented by the genius of Indian people and were derived from pictographs, ideographs, and phonetic signs, the earliest specimens of which are to be found in the Indus Valley inscriptions." An eminent Indian epigraphist, D. C. Sircar (1967, p. 30), envisages that "the Brahmi alphabet seems to have derived from the pre-historic Indus Valley script of a semi-pictographic nature and was popular in the major parts of Bharatvarsha."

The exponents of the second theory, who believed that Brahmi originates from some foreign source, are Otfried Muller, James Prinsep, and E. Senart (Upasak, 1960). They developed and endorsed the theory that Brahmi script had had its source in the Greek script. It was Otfried Muller, who put forward the idea that Greeks introduced the concept of alphabet to Indians when Alexander invaded India. Scholars from the same school of thought also speculated that Greek or Phoenician models imparted the notions to Ashoka's Buddhists to derive their letters (Upasak, 1960). However, these theories have been discarded as they do not support the paleographic and linguistic evidence. William Jones, a philologist of the 19th century, connected the genesis of Brahmi script to the Semitic origin (Taylor, 1883), and thereafter had been supported and followed by innumerable scholars. The views on the Semitic origin theory are roughly divided into three groups, cf. onto those who believe it originates from (1) Phoenician, (2) South-Semitic, and (3) North Semitic. G. Buhler in his book *Indian Paleography* (1904) propounded one of the most influential theories which had received a wide acceptance in Western scholarship for several decades. According to his theory, Brahmi script was derived from an Aramaic alphabet in 8th century B.C.E. He made a comparison between Brahmi and North Semitic alphabets and concluded that twenty-two letters of the Brahmi script were (directly) derived from the North Semitic alphabets, of which some are found in early Phoenician inscriptions (Hartmut Scharfe, 2002; Upasak, 1960). However, Buhler's theories have been challenged and discarded by several Indians as well as Western scholars (for example, see R. Pandey, 1957; Salomon, 1998). Amid the tussle between several theories propounded over the origin of Brahmi script, it is nowadays well accepted that Brahmi alphabets were perfect on phonetic measures.

4 Devanagari

Brahmi-derived scripts are mainly divided into two groups; namely, *Gupta* (northern group) and *Grantha* (southern group). The scripts of Dravidian and a few Austro-Asiatic languages are based on Grantha, while Devanagari and the other scripts of Northern-India are derivatives of the northern group, i.e. Gupta (Patel, 1995). Devanagari, a third

generation offshoot of Brahmi, turned to be the most widely used script in India by the 11th century. In modern India, it coexists with nine other major scripts, including Roman and Perso-Arabic (Vaid & Gupta, 2002). Initially, Devanagari was developed for writing Indo-Aryan classical language Sanskrit, and gradually its use extended to several modern Indo-Aryan languages, like Hindi, Dogri, Nepali, Marathi, Konkani etc. The extension of Devanagari to write other languages, apart from Sanskrit 'conditioned' it with few changes as it was required to represent specific speech-sounds of the newly adopted languages. This conditioning excludes, for instance, the sign *ardha-visarga* or *jihvāmuliya* which means "produced at the root of the tongue" from the modern Devanagari script to write Hindi (Bhat, n.d.; Egenes, 1996). Moreover, a few sounds have been borrowed during the course of development and historical changes. Sounds like /z/, /x/, /ɣ/, /q/ (Perso-Arabic) have been adopted and are being represented by putting a dot beneath the consonant letters **ज** (/dʒ/), **ख** (/kʰ/), **ग** (/g/), and **क** (/k/) to accommodate the contemporary phonological needs of Hindi.

The positioning of alphabets in Devanagari is strictly phonetic, with vowels and diphthongs occurring first and then followed by a sequence of consonants. Vowels, called svára (meaning the reverberation of self) begin with short अ (*a*) followed by its long counterpart आ (*ā*). Explaining the short अ (*a*) vowel, Bhatt (n.d., p. 3) in his paper states that 'the ancient Indian *Śikşa Āçārya*-s (phonetician-s) consider *a* [अ] as the primary sound that appears immediately at the entry point as the pulmonary breath-air enters the vocal tract at the glottis.' In the arrangement of vowels, priority has been given to vowel-length over nasality (*nāsikya*). The arrangement of letters is in accordance to the place of articulation; for example in vowels, the velar अ (*/a/)*, आ (*/ā/)* is followed by the palatal इ (/i/), ई (*/ī/)* and the labial उ (/u/), ऊ (*/ū/*). Other vowels listed thereafter are palatal ए (/e:/) and ऐ (/ɛ:/); velar ओ (/o:/) औ (/œ:/). The ऐ and औ are two velar-palatal and velar-labial diphthongs respectively. The nasal sound has been represented independently as अं (ã). Consonants (*vyañjana*) are positioned from velar to labial where obstruents (*Sparʃa*) occur first, followed by sonorants (*antaħstha*) and sibilants (*Ūʃmāna*) (see Appendix Table 1) (Freund, 2006).

The phonemic units, i.e. consonants and vowels, in Devanagari are represented by two sets of symbols referred (to) as primary and secondary forms. To spell words, the use of these primary and secondary forms is specifically rule-bounded. Mostly, it is the position of a phoneme in a word which determines the rules assigned to both forms. A vowel's primary form is used either when it comes at the beginning of a word or represents a full meaningful unit at its own. The secondary form for vowels, in Hindi, is called *mātra*. These mātras are frequently used in Devanagari after a consonant in a syllable. For example, in Hindi, primary and secondary forms for the vowel /e:/ are 'ए' and 'े'. The primary form is used in the word like एक (/e: k/, one) and the secondary form in the word पेड़ (/pe:ɽ/, tree). Among consonants, the secondary form is used when it comes at the initial or non initial position in a consonant cluster like, पाण्डेय (/pa:ɳɖe:j/, a surname) or पदस्थ (/pəˈdəstʰ/, in position). The primary form is used for

all consonants other than the clusters, occurring at different places in a word. In consonants, the frequency of use of the primary form is much higher than that of vowels, while the secondary form of vowels, i.e. mātra, is more common in writing. Moreover, in Brahmi-derived scripts mātra also represents the unit of time. A small vowel is attributed to the value of one mātra, a long vowel associates with two mātras and a consonant with half of mātra (Patel, 1995).

5 Akshara in Devanagari

Akshara is the orthographic unit of Brahmi scripts. Historically, some researchers have considered it as a precursor of a mora. In North and South Indian languages, the fundamental topographic encoding and the phonological principles are the same but the special visual shape of akshara differs (Vaid & Gupta 2002). Those writing systems that use akshara, like Devanagari, share multiple characteristics with a syllabary but at the same time contain alphabetic features (Nag 2011). Each akshara symbol in a syllabary, represents a syllable. In Hindi, for example, the akshara **चा, चि, चू,** constitute /tʃə:/, /tʃi/, /tʃu:/[1] syllable units. Furthermore, these akshara units can be deconstructed into smaller phonemic units, which show the alphabetic nature of akshara symbols, like, **च् + ◌ा (आ)** (/tʃ/ **+** /ə:/), **च् + ◌ि (इ)** (/tʃ/ + /i/), or **च् + ◌ू (ऊ)** (/tʃ/+/u:/). These individual consonant and vowel sounds within syllable units /tʃə:/, /tʃi/, **/**tʃu:/ resemble English alphabetic sounds, being represented as [ch+ a], [ch + i], [ch+ u].

An akshara can form a nucleus either by itself or with an onset. In case of a coda, it can be formed by itself or can be shifted to the next akshara to assimilate into the onset of the next syllable. There are four main types of symbols in the akshara system; (1) consonants with an intrinsic schwa (Cə), (2) consonants without an inherent short schwa vowel which is marked by a *halant* (C◌्), (3) consonants with other vowels (CV), and (4) consonant clusters (CCV). Consonant clusters can be formed with more than two consonants such as CCCV or CCCVV (Nag, 2011; Patel, 1995). However, Pullum (1971) argues that simply putting together two consonant symbols in Devanagari script does not make a consonant cluster as it does in English. A consonant cluster in Devanagari script is represented by a composite symbol, which is a blend of its component sounds. The visuospatial characteristics of the consonant clusters might have a minimal resemblance to the physical appearance of the letters representing their component sounds. For example, Hindi akshara च represents /tʃ/, क represents /k/, and र represents /r/, so that put together thus चकर would represent /tʃəkər/. However, the accepted Hindi term is चक्र (Wheel) consisting the composite symbol of

[1] Phoneme symbols used are from the IPA, 2002.

क and र sound as क्र (/kr/). The composite symbol of क and र as क्र would now behave like a regular consonant symbol in writing.

Orthographical structure of Devanagari script follows a left-to-right sequencing. It is consonants in the script that follow a strict left-to-right linear order, whereas vowels are positioned non-linearly around them. In writing, vowels act as an adjunct to consonants occurring above, below, or on either side of it, representing the sound sequencing of their spoken forms. However, there are some exceptions where the left-to-right order in writing does not follow the order in which the speech sounds occur. Unlike other vowels, the short vowel /i/, which is represented by the symbols- इ (primary form, placed at the initial position) and ि (secondary or diacritical form, placed at non-initial positions), is attached to the left of the following consonant. Thus, the positioning of a short vowel /i/ creates a discrepancy between written and spoken sequences resulting in Ci (consonant + /i/) in speech and iC (/i/ + consonant) in writing. This can be illustrated with the following example: a word दिल /dɪl/ is written with a vowel diacritic placed before /d/, making the sequence of a medial vowel, an initial consonant, and a consonant (Gaur, 1995; Pullum, 1971). Another distinctive feature of Brahmi-derived scripts, particularly of Devanagari, is that there is a horizontal line going across the top of each word.

6 Nature of Brahmi-derived Scripts: alphabetic, syllabic, alphasyllabic, or something else

The script is a cultural product and its origin and history are placed in a cultural context. Several cultures, in the course of their development, devised their own tools to record their languages. In other cases, already existing writing systems have been adopted to record their languages, or have at least inspired people to create new scripts for their speeches (Upasak, 1960).

I. J. Gelb, one of the pioneering figures in modern times, conducted the most extensive study of the origin and nature of writing systems and general principles of their development. Gelb (1963) in his analysis of writing systems of the world propounded that all scripts, from their origin to full evolution, follow a specific unidirectional stage of development. In his writing, he asserted that no script could skip developmental stages, being logography, syllabography and alphabetography. He writes (1963, p. 201) that “no writing can start with a syllabic or alphabetic stage unless it is borrowed, directly or indirectly, from a system which has gone through all the previous stages.” Further, he states that “there can be no reverse development, i.e., an alphabet can not develop into a syllabary, just as a syllabary can not lead to the creation of logography.” Gelb (1963, p. 144), however, takes a different stand, elsewhere, while describing the origin of Semitic writing. He claims that “the forms are freely invented with new values as found in a large number of writings such as Balti, Brahmi etc.” His

descriptions suggest that he has not addressed the complex identity and developmental process of Brahmi. Contradicting Gelb's categorization, scholars have contended the misguided belief that scripts can only be of three types; logography, syllabary and alphabetic. Similarly, scholars have questioned Gelb's claim of historical evolution and his set stages of development. Daniels (2000; 2002) states that Gelb misleadingly tried to develop an order and symmetry in whatever he explored.

While investigating the unique structure of Hindi writing system, Rimzhim et al. (2014, p. 5) concluded that Hindi orthography is 'functionally predominantly alphabetic'. To claim their argument, they state that "the presence of both full and half forms of vowels puts them orthographically on a par with the full and half forms of consonants respectively... This equivalence is a defining feature of an alphabetic writing system." In response to Rimzhim et al., Share and Daniels (2015) published a paper and listed six reasons why Brahmi-derived scripts should not be called 'alphabetic'. Presenting structural evidence they contended that consonants and vowels are not on a par, as the majority of vowels in Brahmi-derived scripts are not full-sized letters, and are mostly used as *mātras* or left unmarked by occurring inherently. Additionally, in contrast to Greek-derived scripts where consonants and vowels are physically similar, in Devanagari the shape and size of consonants and vowels are not alike, and vowels (in the form of *mātras*) are generally subjoined to consonants, which are larger in size. Further, consonants with a reduced status, i.e. consonants without an inherent short vowel, do not stand equally with a vowel as they maintain a noticeable appearance of the earlier form as a full-sized letter. In other points, too, consonants occur linearly[2] while vowels are positioned nonlinearly, which makes them different from alphabetic systems.

Classification of the Indic writing system is problematic because it does not fit aptly to the traditional typology of writing systems. The specific consonantal syllabic structure with an inherent schwa vowel [Cə] confers a unique identity and sever it from other script categories. Akshara orthographic units, unlike alphabetic scripts, represent sounds at the level of a syllable but at the same time, unlike syllabary scripts, can be broken further into distinct phonemes (see Nag & Sircar, 2008; Nag, 2007). In other words, Indian writing system is syllabic in terms of a syllable (or *akshara*) as a basic graphic unit, but it also reflects a contrary stand to a pure syllabary as discrete sound units of a syllable are identified individualy within the same syllable (Salomon, 1998). Based on these descriptions, a surprising number of scholars have attributed or easily accepted terms such as alphasyllabic, semi-syllabic, sub-syllabic, semi-alphabetic, or neosyllabic when defining the nature of Brahmi and its offshoots. By rejecting the term 'fundamentally alphabetic' in the context of akshara-based scripts, Share and Daniels (2015, p. 6), too, question the term 'alphasyllabic' as they state

[2] Except in the case of /r/, which behaves like a vowel matra at the conjunct position.

> "[W]e argue that they (akshara based scripts) are not fundamentally syllabic. We begin by stating the obvious: in a syllabic script such as Japanese kana, syllable signs cannot be analysed into constituent consonants and vowels. Therefore, the term "alphasyllabic", suggesting that they are somehow a hybrid or mix of the two long-established types, is misleading."

We believe and argue[3] that the Brahmi-derived scripts should not be categorized as scripts with overlapping features of alphabetic and syllabic system as they contain categorization of their own. We suggest a new category named as 'akshara scripts'.

7 Literacy development and teaching of akshara

The role of orthography is to represent speech sounds of a language. Orthographies differ from each other in number of written characters they use to symbolize spoken sounds. Nag (2007) in her work on akshara languages estimated that a reader of akshara languages is required to recollect around 400 or more orthographic units. The learning condition of akshara orthographies became different from alphabetic scripts because of a large number of written symbols used. In Devanagari, for example, orthographic characters uniquely represent single speech sounds in almost all conditions; contrary to English written symbols, which represent more than one speech sound in different environments. Studies suggest (Nag & Sircar, 2008; Nag, 2007) that orthographic learning of akshara system is slower than that of an alphabetic system. Children master letters of alphabetic languages somewhere by the end of their first year of schooling, whereas the akshara (*akshar* means letter) learning continues up to the fourth or fifth grade. Anand (1990) in his study on Hindi found out that fifth grade school children frequently make grapheme errors. To teach akshara symbols, a three-step learning is usually used in classrooms. Children are first taught the consonants with an inherent vowel (Cə), then the consonants with other vowel makers (CV), and finally consonant clusters (CCV). The academically designed Indic script learning is, however, less popular and is being followed differently in places other than academic. Thus, for example, children speaking Kannada[4] get the exposure of CCV symbol unit with the rudimentary Cə symbol unit in their early textbooks, which result in the simultaneous learning of both units (Nag & Sircar, 2008; Nag, 2011).

Several approaches have been adopted or coined to teach Indic scripts; shape-similarity and productive-symbols are two such approaches (Gupta, 2007). For learners of Brahmi-derived scripts, visuospatial characteristics of symbols have always been the issue, rather than sound-symbol correspondence. Like English alphabets 'v' – 'w' or 'b'

[3] David L. Share and Peter T. Daniels (2014) argue the same as 'Brahmi-derived scripts are in a category of their own and merit a unique descriptor.'

[4] Kannada is a prominent language of the Dravidian language family, mainly spoken in the southern part of the Indian subcontinent.

– 'd', which are highly confusing for dyslectic or slow learners, Hindi orthography contains a huge number of symbols with mutual visuospatial characteristics. To solve this issue, Kerslake and Aiyer (1938) wrote a book titled 'Tamil Course for European Schools,' to teach Tamil to students through the shape similarity method. Further, the Central Institute of Indian Languages (CIIL), while developing teaching materials for Indic scripts, categorized written symbols of Devanagari script into eleven categories according to the similarity of their shapes (see Appendix Table 2). Prime objective of such categorization is to make students familiar with possible orthographic details of each written character (see Pattanayak, 1991; Rao, 1978).

In case of the productive symbols method, sounds are clubbed together meticulously to teach basic vocabulary of the language. Mace (1962) in his book developed a new sound sequencing to teach Persian script. For instance, three letters are introduced first – [a], [n], and [b], and then joined in a way to form basic possible words of Persian like ab (آب water), baba (بابا father), an (ان that), nan (نان bread), and banana (بنانا builder). This method has been used to devise script-learning for Indian languages (see Eklavya, 2003; Jayaram, 2008), where symbols are put together in accordance with principles of economy and consistency, to create words immediately. To teach letters of Hindi script, Eklavya (2003) in his book introduces an unconventional sound sequencing. For instance, at one stage, he made the following sound arrangements: क /kə/, ब /bə/, स /sə/, म /mə/, प /pə/, न /nə/, ल /lə/, and a diacritic for /a:/, and at another stage he formed words like न + ल = नल (nə + lə = nəl, tap), क + ल = कल (kə+lə = kəl, tomorrow), and फ + ल = फल (phə+lə = phəl, fruit, result). In Devanagari, when two consonants are put together the inherent vowel at word end is deleted automatically. Instead of following the conventionally phonetic arrangement of letters starting with independent vowels, he focused more on diacritics along with consonants. It is the diacritics in Devanagari, rather than the independent vowel forms, that are used overwhelmingly. Hence, with this approach, children generally learn the complete word at a time, while they also get familiar with the grapheme-phoneme mapping in Hindi language.

To spell words correctly, it is essential for a child to master the skill of connecting individual phonemes with corresponding orthographic units. The process of spelling makes a child aware of the units of meaning (morphemes), and the grapho-phonic knowledge of a language (Weeks, Brooks, & Everatt, 2002; Westwood, 2005). However, strong impact of the phonological domain has been observed over the orthographic domain, which suggests that both the domains are not on a par. Unique dialectical sounds in a child's spoken language, varying from the standard spoken and written sounds, are difficult to spell as discrepancy emerges between the standard phonological unit and the one that a child has inherited though dialectical sounds. In Kannada, for instance, Nag et al. (2010) found out that the glottal /h/ sound is difficult for children to spell correctly as an inconsistency occurs between mapping the standard spoken and the written form by a specific dialect feature.

Share and Daniels (2015, p. 11) suggest that if Brahmi-derived scripts are considered as alphabetic in nature, then all scientific advancement in the field of English literacy learning can be implemented on them. On the contrary, if Indic scripts merit a unique identity based on the features they show as an orthography, "instruction will need to focus on more psycholinguistically accessible supra-phonemic units."

8 Conclusion

The nature of Brahmi-derived scripts, particularly of Devanagari, is often termed as alphasyllabic. Some researchers believe that the alphasyllabic attribution to the Indic scripts can be found in overlapping features of alphabetic and syllabary writing systems. On the contrary, there are some researchers who assert that the Indic scripts are neither fundamentally alphabetic nor fundamentally syllabic. It is because of the strong influence on the Indian academia of the identity and methods that originate in the Western academia, the attributions like alphasyllabic or semi-alphabetic for Bhrami-derived scripts are readily accepted. It is essential to understand that the complex architecture of Devanagari script superficially presents some alignment with the alphabetic as well as syllabic properties. However, the fundamental property of akshara units of the Devanagari script is distinctive in its nature. Thus, the Devanagari script should not be termed as any of the two types and demand a unique descriptor.

References

Anand, V. (1990). *Hindi spelling: Errors and remedies*. Bhavana Prakashan.

Anderson, J. M., & Ewen, C. J. (1987). *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Basham, A. L. (1967). *The Wonder that was India.* New York: Mac.

Bhat, R. (n.d.). *Devanagari*. Retrieved from https://www.academia.edu/11307177/DEVANAGARI

Bright, W. (1996). The devanagari script. In P. T. Daniels & W. Bright (Eds.), *The World's Writing Systems* (pp. 384-390). Oxford University Press.

Bühler, G. (1904). Indian Paleography, from about BC 350 to about AD 1300. *Bombay education society's Press*, Vol. 1.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77-111. https://doi.org/10.1016/S0010-0277(03)00164-1

Egenes, T. (1996). *Introduction to Sanskrit*. Motilal Banarsidass Publ. Retrieved from http://abhidharma.ru/A/Raznoe/Yaz/Ind/0001.pdf

Eklavya (2003). *Padho Likho Maza Karo: Part 1.* Bhopal: Eklavya.

Freund, P. F. (2006). *Vedic literature reading curriculum*. Available from ProQuest Dissertations & Theses Global.

Gaur, A. (1995). Scripts and writing systems: A historical perspective. In I. Taylor & D. R. Olson (Eds.), *Scripts and literacy* (pp. 19–30). Dordrecht: Kluwer Academic.

Gelb, I. J. (1963). *A study of writing* (Revised edition). Chicago, IL: University of Chicago Press. Retrieved from https://oi.uchicago.edu/sites/oi.uchicago.edu/files/uploads/shared/docs/study_writng.pdf

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of phonetics, 29*(4), 383-406. https://doi.org/10.1006/jpho.2001.0147

Gupta, R. (2008). Initial literacy in Devanagari: What matters to learners. *South Asia Pedagogy and Technology 1*. University of Chicago. Retrieved from http://hdl.handle.net/11417/1140

Jayaram, K. (2008). Early Literacy Project-Explorations and Reflections, Part 2: Interventions in Hindi Classrooms. *Contemporary Education Dialogue*, *5*(2), 175-212. https://doi.org/10.1177/0973184913411166

Kapoor, S. (2002). *The Indian Encyclopedia; Biographical, Historical, Religious, Administrative, Ethnological, Commercial and Scientific*. New Delhi: Cosmos Publishing

Kaye, J., Lowenstamm, J., & Vergnaud, J.-R. (1985). The internal structure of phonological elements: a theory of charm and government. *Phonology*, *2*(1), 305-328. https://doi.org/10.1017/S0952675700000476

Kerslake, P. C., & Aiyar, C.R. Narayanaswami. (1938). *Tamil Course for European Schools*. Madras: The Christian Literature Society for India.

Mace, J. (1962). *Teach yourself modern Persian*. London: English Universities Press.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic inquiry*, *12*(3), 373-418. Retrieved from http://www.jstor.org/stable/4178229

Mukherjee, P. K., Nema, N. K., Venkatesh, P., & Debnath, P. K. (2012). Changing scenario for promotion and development of Ayurveda-way forward. *Journal of Ethnopharmacology, 143*(2), 424-434. https://doi.org/10.1016/j.jep.2012.07.036

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading*, *30*(1), 7-22. https://doi.org/10.1111/j.1467-9817.2006.00329.x

Nag, S. (2011) The akshara languages: what do they tell us about children's literacy learning? *Language-cognition: state of the art*, 291-310. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.5579&rep=rep1&type=pdf

Nag, S., & Sircar, S. (2008). *Learning to read in Bengali: Report of a survey in five Kolkata primary schools*. Bangalore, India: The Promise Foundation.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research*, *2*(1), 41-52. https://doi.org/10.1093/wsr/wsq001

Ojha, G. H. (1959). *Prachina Bharatiya Lipimala*.

Pandey, P. (2007). Phonology-orthography interface in Devanāgarī for Hindi. *Written Language & Literacy*, *10*(2), 139-156. https://www.jnu.ac.in/Faculty/pkspandey/papers/final_proofs-_devnagari_script.pdf

Pandey, R. (1957). *Indian palaeography* (Vol. 1). Motilal Banarasi Das.

Patel, P. G. (1995). Brahmi scripts, orthographic units and reading acquisition. *Scripts and literacy: Reading and learning to read alphabets, syllabaries, and characters*, *7*, 265. https://link.springer.com/chapter/10.1007/978-94-011-1162-1_17

Pattanayak, D. P. (1991). *Language, education, and culture* (Vol. 46). Central Institute of Indian Languages.

Pullum, G. K. (1971). Indian scripts and the teacher of English. *ELT Journal*, *25*(3), 278-284. https://doi.org/10.1093/elt/XXV.3.278

Rao, G. S. (1979). *Literacy methodology: papers presented at the National Seminar on Methodology of Literacy Material Preparation at Mysore in 1978* (Vol. 3). Central Institute of Indian Languages.

Rimzhim, A., Katz, L., & Fowler, C. A. (2014). Brāhmī-derived orthographies are typologically Āksharik but functionally predominantly alphabetic. *Writing Systems Research*, *6*(1), 41-53. https://doi.org/10.1080/17586801.2013.855618

Roop, H. D. (1972). *An introduction to the Burmese writing system*. New Haven: Yale University Press.

Ruhlen, M. (1991). *A guide to the world's languages: Classification* (Vol. 1). Stanford, Calif: Stanford University Press.

Salomon, R. (1998). *Indian epigraphy: A guide to the study of inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan languages*. New York: Oxford University Press.

Scharfe, H. (1977). *Grammatical literature* (Vol. 2). Wiesbaden: Otto Harrassowitz Verlag.

Scharfe, H. (2002). Kharosti and brahmi. *Journal-american oriental society*, *122*(2), 391-393. https://www.jstor.org/stable/pdf/3087634.pdf

Share, D. L., & Daniels, P. T. (2016). Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler (2014). *Writing Systems Research, 8*(1), 17-31. https://doi.org/10.1080/17586801.2015.1016395

Sircar, D. C. (1957). *Inscriptions of Asoka*. Publications Division Ministry of Information & Broadcasting.

Taylor, I. (1883). *The alphabet: an account of the origin and development of letters* (Vol. 2). K. Paul, Trench & Company.

Upasak, C. S. (1960). *The history and palaeography of Mauryan Brāhmī script*. Nava Nālandā Mahāvihāra.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language*, *81*(1), 679-690. https://dspacepre1.library.tamu.edu/handle/1969.1/158735

Verma, T. P. (1971). *The palaeography of Brāhmī script in north India, from c. 236 BC to c. 200 AD*. Siddharth Prakashan.

Weeks, S., Brooks, P., & Everatt, J. (2002). Differences in learning to spell: Relationships between cognitive profiles and learning responses to teaching methods. *Educational and Child Psychology*, *19*(4), 47-62.

Westwood, P. (2005). *Spelling: Approaches to teaching and assessment*. Aust Council for Ed Research.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, *131*(1), 3. doi:10.1037/0033-2909.131.1.3

APPENDIX

Table 3: Phonological inventory of an Indic script

Vowels	Primary Vowels	**अ** a	**आ** ā	**इ** i	**ई** ī	**उ** u	**ऊ** ū	**ऋ** ɻ̩
	Secondary Vowels	**ए** e	**ऐ** ai	**ओ** o	**औ** au	**अं** ə̃	**अः** ah	

Consonants		Voiceless Plosives		Voiced Plosives		Nasals
		Unaspirated	Aspirated	Unaspirated	Aspirated	
	Velar	**क** kə	**ख** k^hə	**ग** gə	**घ** g^hə	**ङ** ŋə
	Palatal	**च** tʃə	**छ** $tʃ^h$ə	**ज** dʒə	**झ** $dʒ^h$ə	**ञ** ɲə
	Retroflex	**ट** ʈə	**ठ** $ʈ^h$ə	**ड** ɖə	**ढ** ɖə	**ण** ɳə
	Dental	**त** t̪ə	**थ** $t̪^h$ə	**द** d̪ə	**ध** $d̪^h$ə	**न** nə
	Labial	**प** pə	**फ** p^hə	**ब** bə	**भ** b^hə	**म** mə

Semi-Vowels	**य** jə	**र** ɾə	**ल** lə	**व** ʋə
Sibilants	**श** ʃə	**ष** ʂə	**स** sə	**ह** ɦə

Table 4: Symbols based on shape similarity

Group	Devanagari Symbols	Pronunciation
1	व, क, ब	ʋə, kə, bə
2	ग, म, भ, झ	gə, mə, b^hə, $dʒ^h$ə
3	र , स , ख , ए , ऐ , श	ɾə , sə , k^hə , e: , ai , ʃə
4	ण, प, ष, फ	ɳə, pə, ʂə, p^hə
5	त, न, ल	t̪ə , nə , lə
6	ट , ठ , ढ , ढ॰ , द	ʈə , $ʈ^h$ə , $ɖ^h$ə , ɽ^{h}ə, d
7	ड , ड़ , इ , ई , ह	ɖə , ɽə , i , ī , ɦə
8	घ, ध, छ	g^hə, $d̪^h$ə , $tʃ^h$ə
9	च, ज	tʃə, dʒə
10	उ , ऊ , अ , आ , ओ , औ	ʊ , ū , ə , ā , o: , ɔ:
11	य , थ	jə , $t̪^h$ə

IMAGE OF JAPAN AMONG SLOVENES: BORROWED WORDS OF JAPANESE ORIGIN IN SLOVENE

Chikako SHIGEMORI BUČAR
University of Ljubljana, Slovenia
chikako.bucar@guest.arnes.si

Abstract

This paper presents the process and mechanism of borrowing from Japanese into Slovene. Japan and Slovenia are geographically and culturally quite distant, and the two languages are genealogically not related. Between such two languages, not many borrowings are expected, but there is a certain amount of borrowed words of Japanese origin in today's Slovene. The focus of this paper is on the words of Japanese origin that are well integrated in today's Slovene. Firstly, the process of borrowing is analysed: there are three main phases for successful borrowing from Japanese into Slovene, but during the process, some obstacles may hinder the completion of this process, so that further creative use of some borrowed words in the Slovene environment cannot be expected. The second part of this paper will closely look at the loanwords of Japanese origin which are already recorded as headwords in today's dictionaries of Slovene. The loanwords are analysed in relation to the borrowing process and adjustments, their semantic fields, and wherever possible, their diachronic changes in use, and other specifics. At the end, the image of Japan seen through the borrowing process and consolidated loanwords is summarized, and possible development of borrowing in the near future is predicted.

Keywords: loanwords; Slovene; Japanese; borrowing; derivation; number and gender

Povzetek

Članek v prvem delu obravnava postopek in mehanizem izposojanja iz tujega jezika v slovenščino, zlasti v primeru izposojanja iz jezika, ki je kulturno in jezikovno oddaljen, kakor je japonski. Drugi del članka je pregled izposojenk, ki imajo svoj izvor v japonkem jeziku in so danes že gesla v slovarju slovenskega jezika. Sledi analiza izposojenk glede na postopek vključevanja v slovensko besedišče in semantično polje, v katerem deluje posamezna beseda. Analiza vključuje, kolikor je možno, spremembe in rabe posamezne izposojenke skozi zgodovino ter druge značilnosti. Na koncu je povzetek današnjega stanja izposojenk in njihove rabe, ki ponuja določeno podobo japonske kulture v slovenski družbi.

Ključne besede: izposojenke; slovenščina; japonščina; izposojanje; izpeljava; število in spol

Acta Linguistica Asiatica, 9(1), 2019.
ISSN: 2232-3317, http://revije.ff.uni-lj.si/ala/
DOI: 10.4312/ala.9.1.75-88

1 Introduction

Japan and Slovenia are geographically and culturally quite distant, and the two languages, Japanese and Slovene, are genealogically not related. Between such two languages, not many borrowings are expected, and yet, there have been indirect and direct contacts of the two cultures and nations, particularly after the end of Tokugawa era in 1868. Therefore, borrowings of originally Japanese words do exist in contemporary Slovene. Most of the borrowings from Japanese into Slovene occurred in the 20th and 21st century. From the linguistic point of view, the borrowing mechanism is quite interesting in the case of Japanese words into Slovene, since the phonetic and lexico-grammatical differences between the two languages demand various adjustments for the borrowed words to become loanwords in the Slovene local environment.

The focus of this paper is on the words of Japanese origin that are well integrated in today's Slovene. Firstly, the process of borrowing is analysed: there are three main phases for successful borrowing from Japanese into Slovene, but during the process, some obstacles may hinder the completion of this process, so that further creative use of some borrowed words in the Slovene environment cannot be expected. The second part of this paper will closely look at the loanwords of Japanese origin which are already recorded as headwords in today's dictionaries of Slovene. The loanwords are analysed in relation to the borrowing process and adjustments, their semantic fields, and wherever possible, their diachronic changes in use, and other specifics. At the end, the image of Japan seen through the borrowing process and consolidated loanwords is summarized, and possible development of borrowing in the near future is predicted.

2 Process and mechanism of borrowing

Figure 1 is a schematic summary of the process of borrowing from Japanese into Slovene. There are three main phases for each word, usually a noun, to be accepted as a loanword into Slovene (a., b. and c. in Figure 1). If a borrowed word reaches the last phase (c), we can say that it has fully become a loanword in Slovene, i.e. it is used freely and creatively in the Slovene context.

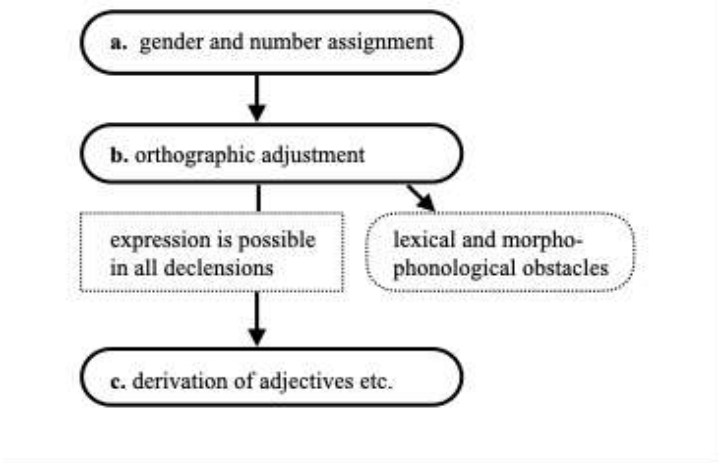


Figure 1: Three phases for successful borrowing

2.1 Gender and number assignment

The first phase (a.) of borrowing is the assignment of gender and number to the new borrowed word. The Japanese langauge has no category of number or of gender. On the other hand, since the categories of number and gender exist in Slovene, every Japanese noun to be used in the context of Slovene must be categorized into one of the numbers (singular, dual or plural) and genders (masculine, feminine or neutral). The gender and number of each new word is usually decided according to the phonetic form of the word. Since most of the Japanese syllables are open and end in one of the five vowels *a*, *e*, *i*, *o* or *u*, the gender is assigned according to the ending vowel, i.e. *-a* as feminine, *-e, -i, -o* and *-u* as masculine. The assignment of number is basically singular. Words with the moraic nasal /N/ (*-n*) at the word end is also categorized as masculine (See Table 1)[1]. Exceptional cases are given with examples on the right side of Table 1: under ② <5> when a word with a final vowel *-e* is interpreted as a plurale tantum (example: *karaoke*), and under ③ <3> , <6> and <11> when nouns with a final vowel *-a, -e* or *-o* are interpreted as masculine with slightly different ways of declension.[2]

Most of the gender and number assignment is done according to the morphological rules under ① in Table 1, that is, originally Japanese words are usually categorized as feminine or masculine in gender, and in most cases singular in number.

[1] This table was made to present all theoretically possible solutions of the gender and number assignment, from <1> to <13> in the table, but for <2> and <8> no example exists. For detailed explanation, please also refer to Shigemori Bučar (2011).

[2] ③ <3> is a case when a noun with a final vowel *-a* is interpreted as masculine and assigned the second masculine inflection in Slovene. ③ <6> and <11> are cases when the last vowel of a longer noun is observed as a non-voiced or unaccented *-e* or *-o* and declined differently from the cases under ① <4> and <10>. For details, please also refer to Shigemori Bučar (2011).

Table 1: Possible solutions of gender and number assignments to nouns of Japanese origin in Slovene (Shigemori Bučar (2011, p. 251))

Word ending in Jap.	①		②		③	
	by morphological rules; singular	example*	plural interpretation	example	additional rules and interpretations	example
-a	f.sg. <1> [+anim] [–anim]	*gejša -e* *ikebana -e*	n.pl. <2>	/	m.sg. <3> [+anim]	*jakuza -e*
-e	m.sg.(-j-)<4>	*anime -ja*	f.pl. <5>	*karaoke* *karaok* *šitake šitak*	m.sg. <6> [+anim]	*kamikaze -a*
-i	m.sg.(-j-)<7>	*cunami -ja*		/	-Vi>-Vj m.sg. <9>	*samuraj -a*
-o	m.sg.(-j-)<10>	*go -ja*			m.sg. <11>	*kimono -a*
-u	m.sg.(-j-)<12>	*tofu -ja*				
-n	m.sg.(-j-)<13>	*šogun -a*				

* Each example is shown in its nominative and genitive form

2.2 Orthographic adjustment

The second phase of the borrowing from Japanese into Slovene is the orthographic adjustment (b. in Figure 1). There are some examples of words that end in the vowel sequence *-ai* in Japanese, for which the *-i* at the end is replaced with a *-j* in the Slovene orthography (under ③ <9> in Table 1, example: *samuraj*). This replacement, which may be called "slovenization" in writing and speaking, occurs most probably because of similar Slovene words with the ending *-aj* /-ai/ (e.g. *čuvaj* 'custodian'; *papagaj* 'parrot'), and because the borrowed words may then be declined more easily in the same manner as the existing Slovene words, i.e. *čuvaj* [nom], *čuvaja* [gen], *čuvaju* [dat]; *samuraj, samuraja samuraju*.

Today's most frequent and worldwide way to romanize Japanese is the Hepburn romanization developed in the late 19th century. On the other hand, the modern Slovene uses a set of Latin alphabet called "gajica" (Gaj's alphabet) devised by Croatian linguist Ljudevit Gaj in 1835, based on Jan Hus's Czech alphabet. The principle of Gaj's alphabet is that every sound should have only one letter. Japanese consonants for which digraphs are used in the Hepburn system, i.e. *ch*, *ts* and *sh*, are rewritten in the course of borrowing by one letter in Slovene, in case of *ts* by *c*, in case of *ch* and *sh* with a caron (or a hachek=inverted circumflex), *č* and *š* (examples: *tsunami* → *cunami; matcha* → *mača*[3]*; shiitake* →*šitake*[4]). The sound spelled *j* /ʒ/ in the Hepburn system is

[3] Here is an additional adjustment with the glottal stop, described in the followig paragraph.

[4] Here, too, is an additional adjustment with the long vowel, described in the following paragraph.

written with the corresponding *ž* or with the digraph *dž* in Slovene. The letter *y* is not used in Slovene (except for foreign proper nouns), and the sound represented by *y* in the Hepburn system is spelled with *j* in Slovene (example: *yakuza* → *jakuza*).

Other disputable points in orthography and phonology are the opposition of long and short vowels: this distinction in Japanese is not expressed in everyday press in English, German etc., and since Slovene also has no distinction of long and short vowels, the originally Japanese loanwords lose this distinction in the Slovene environment (i.e. *budô* → *budo*). A similar problem exists with the glottal stop usually romanized in the Hepburn system as a double consonant. The Slovene authority is of the opinion that a double consonant of foreign origin must be rewritten with a single consonant. Therefore, loanwords are spelled (and pronounced) *mača* for *matcha*, *šitake* for *shiitake*.

Orthography is closely connected to phonology. Some cases at this phase of borrowing changes the phonetic form of the borrowed word, and its consequence is the audial and morphological 'deformation' (from the standpoint of the original language) or 'integration' (from the standpoint of the new langauge into which the word is being borrowed). The "slovenized" forms may sometimes trigger association with similar lexical items in Slovene, for example, *sushi* → *suši,* which is similar to the Slovene verb *sušiti* 'to dry'. The nominalization of this Slovene verb leads to the noun *suša* "draught", of which the dative or locative form would be *suši*, which may confuse some users of the Slovene language when they encounter a new loanword *suši* to indicate a Japanese dish.

2.3 Lexical and morphophonological obstacles

If a borrowed noun cannot be used freely in the new environment due to some lexical or morphophonological conditions, it leads to the question of whether or not the word will gain its position as a loanword. The following conditions may decide if a word can completely integrate into the Slovene language or not:

A. lexical space
B. morphophonological clearness
C. possiblilty for further derivation

The first condition above, A, has been explained in the previous section with the example *suši*. If the place to be accepted, in this case the Slovene lexicon, is "crowded", it is difficult for the new word to gain its position as a part of the lexicon. There are cases when a new word may be too short or too long, or phonologically strange or impossible to be used in the new environment (condition B above). *Ukiyo-e* is such an example. According to the corpus Gigafida[5], this word is used only in this form (in

[5] The Gigafida corpus is an extensive collection of Slovene text of various genres, from daily newspapers, magazines, all kinds of books (fiction, non-fiction, textbooks), web pages, transcriptions

nominative) without any declined case forms. In fact, the morpheme *e*, "picture" in Japanese, is very short and there is already a headword "e" in Slovene dictionaries (an interjection). Besides, this short borrowed word is used in the Slovene environment only as a terminology in art history, in combination with the concept *ukiyo*. Though there have been exhibitions of the Japanese woodcut prints in Sovenia, and this genre of art has become quite popular in Europe and also in Slovenia, the word is not found in the existing Slovene dictionaries as a loanword. For a comparison, a similar word composition *yamato-e*, in the partly slovenized form *jamato-e*, was used in 1999, according to the corpus Gigafida, but much less often than *ukiyo-e*.

2.4 Derivation of adjectives and further development

On the other hand, flexibility and creativity of the users of Slovene can be felt in cases when new words are derived from the loanwords of Japanese origin (condition C above). The derivation is possible when and only when a loanword is felt quite integrated and used frequently in the new Slovene environment, so that there is the urge among the users to derive new words according to the Slovene grammatical rules. In case of the present research, there are some adjectives derived from loanwords of Japanese origin (phase c. in Figure 1):

haiku → haikujevski/haikujski	"of haiku, in the manner of haiku"
samuraj → samurajski/samurajev	"of samurai, samurai's"
kamikaze → kamikazov/kamikazin	"kamikaze's"

Both words of the first example are mentioned in the present Slovene etymological dictionary as a subentry under the headword *haiku*, and it seems that the two forms are competing with each other. The second and third examples are found in the dictionary of Slovene orthography, *ePravopis Slovenian Normative Guide*[6], both as headwords.

According to the existing dictionary entries, some other loanwords from Japanese are used as adjectives in their noun forms, without deriving a new adjective. In the traditional Slovene grammar, such adjectival use of a noun, usually placed before the head noun of the phrase, is not recommended, but such expressions are gaining ground in Slovene due to the recent influence from the English language, e.g. *shiatsu*: *V Švici je opravila študij shiatsu terapije in zdrave prehrane. (=In Switzerland she completed the study of shiatsu therapy and healthy diet.)*. Similarly, there is an adjectival use of the

of parliamentary debates and similar. It contains almost 1.2 billion words, or exactly 1,187,002,502 words. (accessed 16.9.2018: http://eng.slovenscina.eu/korpusi/gigafida)

[6] ePravopis Slovenski pravopis 2014 – 2017, Inštitut za slovenski jezik Frana Ramovša ZRC SAZU

noun *sumo*: *sumo borec* (*sumo wrestler* in English), but in this case, there is also an entry in the dictionary as one word, i.e. a further derivation in Slovene: *súmobórec*.[7]

3 Loanwords of Japanese origin in Slovene dictionaries

The abbreviated label "jap", meaning "*japonščina, japonski* (=Japanese language, Japanese)", was used to search for loanwords from Japanese which are recorded as headwords in today's Slovene dictionaries. This was done on the internet portal Fran:

> The *Fran* portal brings together dictionaries, Slovenian linguistic sources and portals that took shape or are currently under development at the *Fran Ramovš Institute of the Slovenian Language* at the Scientific Research Centre of the Slovenian Academy of Sciences and Arts (ZRC SAZU), as well as dictionaries that have undergone the process of retrodigitization within the Institute's framework.[8]

Total 34 headwords were found with the label "jap" in the following three dictionaries:

1. eSSKJ Slovar slovenskega knjižnega jezika 2016 –2017
   (=Dictionary of the Slovenian Standard Language, 3rd Edition)
2. Slovenski etimološki slovar[3] 2015
   (=Slovenian Etymological Dictionary, 3rd edition)
3. Slovar novejšega besedja slovenskega jezika 2014
   (=Dictionary of New Slovenian Words)

Table 2 is the list of these 34 loanwords labeled "jap" in the three dictionaries. In the table, the headwords (in nominative and genetive forms, as it is usually the case in dictionaries) are listed in alphabetical order, with their assigned gender, and the name of the dictionary in which it is mentioned, the original word in today's Japanese, and other data and comments. As can be seen, there are 35 headwords listed, but numbers 6 and 13 are one and the same concept, only spelled differently in two different dictionaries (*džudo* and *judo*).

There are certainly much more borrowings from Japanese in today's Slovene. For example, the list of Slovene words of Japanese origin in Wikipedia has currently 153 entries.[9] Some examples in this paper were also taken from elsewhere (corpus Gigafida) to illustrate the process of borrowing and obstacles: *ukiyo-e* etc.

---

[7] Dictionary of the Slovenian Standard Language, 3rd Edition

[8] https://fran.si/o-portalu

[9] https://sl.wikipedia.org/wiki/Seznam_slovenskih_besed_japonskega_izvora, most recent changes on 5. January, 2018. (accessed 16. 9. 2018)

Here, the scope is within the limit of the label “jap” which indicates that the words are recognized by Slovene lexicographers as loanwords from Japanese, and they exist and have been used in the Slovenian context at least for some time. The 34 loanwords are analyzed below in relation to the assignment of gender and number, phonological adjustment, etymology and semantic fields.

Table 2: Loanwords of Japanese origin in Slovene

No.	Headword	Gen.	Dictionary	Japanese	Other information	Sem. field
1	aikido aikida	m	eSSKJ, SNB	合気道 (あいきどう)	Eng. aikido	《2》
2	animé -êja	m	SNB	アニメ	Eng. anime	《4》
3	búto -a	m	SNB	舞踏 (ぶとう)	also *butoh -a* Eng. butoh	《4》
4	cunámi -ja	m	SNB	津波 (つなみ)	also *tsunami -ja* (SSKJ) Eng. tsunami	《6》
5	džîudžîcu -a	m	Etym.	柔術 (じゅうじゅつ)	Germ. Jiu-Jitsu or Eng. jiu-jitsu, ju-jitsu	《2》
6	džûdo -a	m	Etym.	柔道 (じゅうどう)	Germ. Judo, Eng. judo	《2》
7	gẹ̑jša -e	f	Etym.	芸者 (げいしゃ)	Germ. Geisha and Eng. geisha	《4》
8	gọ̑ -ja	m	Etym.	碁 (ご)	Germ. Go	《5》
9	haiku -ja	m	SNB	俳句 (はいく)	*haiku*[2] – adjectival use *haikujevski, haikujski* Eng. haiku	《4》
10	harakîri -ja	m	Etym.	腹切り (はらきり)	Germ. Harakiri	《3》
11	ikebâna -e	f	Etym.	生け花 (いけばな)	Germ. Ikebana	《4》
12	jakúza -e	m	SNB	やくざ	Eng. yakuza	
13	judo juda	m	eSSKJ	柔道 (じゅうどう)	[júdo] and [džûdo] Germ. Judo, Eng. judo	《2》
14	káki[1] -ja	m	Etym.	柿 (かき)	*kaki*[2] Khaki It. cachi and New Latin (Diospyrus) kaki	《1》
15	kamikâze -e/-a	m	Etym.	神風 (かみかぜ)	Germ. Kamikaze and Eng. kamikaze	《6》
16	karaóke -ók karaọ̑ke -ọ̑k	f pl.	SNB, Etym.	カラオケ	Eng. karaoke	《5》
17	karatẹ̑ -ja	m	Etym.	空手 (からて)	Germ. Karate	《2》
18	katána -e	f	SNB	刀 (かたな)	Eng. katana, Germ. Katana	《3》

No.	Headword	Gen.	Dictionary	Japanese	Other information	Sem. field
19	kimǫ̑no -a	m	Etym.	着物 (きもの)	Germ. Kimono	
20	mánga -e	f	SNB	漫画 (まんが)	Eng. manga	《4》
21	náši -ja	m	SNB	梨 (なし)	also *nashi -ja* Eng. nashi	《1》
22	níndža -e	m	SNB	忍者 (にんじゃ)	Eng. ninja	《3》
23	reiki -ja	m	SNB	霊気 (れいき)	in alternative medicine Eng. reiki, Germ. Reiki	《7》
24	rikša rikše rîkša -e	f	eSSKJ, Etym.	力車 (りきしゃ)	Germ. Rikscha, Jap. jinrikisha	
25	samurâj -a	m	Etym.	侍 (さむらい)	Germ. Samurai, Fr. samurai, samouraï,	《3》
26	sêjtan -a	m	SNB	せいたん	also *seitan -a* Eng. seitan, Germ. Seitan	《1》
27	shiatsu[1]—/-ja	m	SNB	指圧 (しあつ)	[šijácu] also *šiácu —/-ja shiatsu*[2] adjectival use Eng. shiatsu, Germ. Shiatsu	《7》
28	sǭja -e	f	Etym.	—	German Soja, Dutch soja	《1》
29	sudóku -ja	m	SNB	数独 (すうどく)	Eng. sudoku	《5》
30	súmo[1] -a	m	SNB	相撲 (すもう)	Eng. sumo, German Sumo *súmo*[2] adjectival use	《2》
31	súši -ja	m	SNB	寿司 (すし)	Eng. sushi	《1》
32	šitáka -e	f	SNB	椎茸 (しいたけ)	usually plural also *šitáke* — Eng. shiitake	《1》
33	tajkún -a	m	SNB	大君 (たいくん)	Eng. tycoon Ch. taijun	
34	tamagóči -ja	m	SNB	たまごっち	Eng Tamagotchi	《5》
35	tofú -ja	m	SNB	豆腐 (とうふ)	Eng. tofu Ch. doufu	《1》

3.1 Gender and number assignment

33 headwords out of 34 are categorized as singular. There is one exceptional case, *karaoke,* to which plural number is assigned. The reason seems to be the semantic analogy of Slovene speakers. In Slovenia, there are several place names with the feminine plural ending *-e,* e.g. Jesenice, Medvode etc. Since *karaoke* is perceived as a certain place where singing takes place, the word was accepted to Slovene as a plurale tantum. The word started to appear in the middle of 1990s in Slovene texts, and it is

now regularly used as a plural noun. In all other cases, words ending in the vowel *-a* are feminine, words with endings *-e*, *-i*, *-o* and *-u* are masculine singular. In case of *shiitake*, the original form for a sort of mushrooms was understood by the Slovene users as plural, and the singular form was created with the ending vowel *-a, šitaka,* which is now the headword in the dictionary. The word for 'mushroom' in Slovene is *goba* and has the vowel ending *-a*, which may have triggered this creation. It is interesting to observe that all words with more than two syllables with the ending vowel *-o* has the Slovene genitive ending *-a,* while the one-syllable word *go* is declined with the extension for foreign words *-j-*, therefore, when declined, *go* [nom]*, goja* [gen]*, goju* [dat] and so on. It is to be noted that among these words which end in *-o, buto, judo* and *sumo* originally had a long vowel *-ô* in Japanese. Particularly such short words may cause difficulties in communication when they are used in other declined forms (*buta* [gen]*, butu* [dat]*, z butom* [inst]).

3.2 Phonological adjustment

All long vowels *ô* and *û* in original Japanese lose their length when accepted as loanwords in Slovene. However, it is not the case with the long vowel *ê*. Since they are written in the Japanese orthography as a sequence of two syllables, i.e. *-e* followed by an *i*, they have kept the form of two consecutive vowels *e* and *i* in Latin alphabet. Some of them went through the orthography adjustment in the same manner as the sequence *-a* followed by an *i* which was mentioned in section 1.2 above, *ai* → *aj*. Therefore, *-ei* → *-ej: geisha* → *gejša*, *seitan* → *sejtan*.

While Japanese is a pitch-accent language, Slovene is a stress-accent language.[10] Generally, the stress falls on the second syllable from the last in Slovene. All words on the list in Table 2 abide by this Slovene rule. It is interesting to observe that the two-syllable words *kaki* "persimmon" and *naši* "Japanese pear" also get the stress accent on the second syllable from the last, i.e. the first syllable in the word, so if they are heard by users of the standard kantô dialect Japanese, they would suggest another meaning, *kaki* as "oyster" 牡蠣 and *naši* as "none" 無し, respectively.

Some of the spellings and pronunciation of the loanwords reveal the history of these words/concepts outside Japan: *džiudžicu* (*jûjutsu*) and *džudo* (*jûdô*). For the former case, the German and English variants were obviously from the time before the Japanese reform for modern *kana* usage in 1946. The *Fran* portal also has a section of language and terminological counselling, and a question of how to spell "jujutsu" in Slovene was answered in detail in August 2015.[11] However, the generally circulated version, *džiudžicu,* is still the headword in the etymological dictionary. In the latter case,

[10] Some dialects in Slovenia retain the tonal accent system even today, but two thirds of Slovenia do not practice the tonal accent, and "standard" Slovene is said to be a stress-accent language.

[11] https://svetovalnica.zrc-sazu.si/topic/993/kako-pisati-ime-športa-jujutsu (accessed 16. 9. 2018)

the newer version *judo* is regularly used. The sport is very popular among Slovene children and adults. The widely practiced pronunciation is /judo/ ユド and outweighs the variant /džudo/ ジュド.

3.3 Etymology

The dictionary entries contain detailed etymological information, i.e. from which languages these loanwords were taken by Slovene users. None of the 34 loanwords were borrowed directly from Japanese, but most of them through German (6 words), English (14 words) or both German and English (9 words). Other than these, *kaki* (the fruit persimmon) came through Italian and New Latin, *samuraj* through German and French, *taikun* and *tofu* are said to be taken over from English with the help of Chinese forms, *taijun* and *doufu.* In the case of *soja* (*soya* in English)*,* the etymological dictionary says that the loanword came through English and Dutch and the original form in Japanese is *shôyu*. This may be one of the oldest loanwords in Slovene with their origin in Japan, since the Dutch were present in Japan in the beginning of 17th century. It may be that the meaning has shifted in the lapse of time from "soy sauce" to "soy beans", since beans are called *daizu* in Japanese. Another interesting word on the list is *sejtan* or *seitan* in English*.* This word is not commonly known in today's Japan. According to sources, *seitan* is another name for gluten meat*.* The word was coined in 1961 by George Ôsawa (in Japan Yukikazu Sakurazawa), a Japanese advocate of the macrobiotic diet. In Japanese, the *katakana* naming (a loanword from the West) "guruten mîto" is more usual. It is not known exactly how *seitan* should be written in Japanese. Possible variations are 正蛋, 生蛋, 製蛋.[12]

3.4 Semantic fields

The 34 loanwords of Japanese origin were classified into groups according to their semantics. The numbers on the far right in Table 2 show the following semantic classes:

《1》	Cooking and food	7 words
《2》	Sports and martial arts	6 words
《3》	Samurai culture	4 words
《4》	Art	6 words
《5》	Games and toys	4 words
《6》	Climate	2 words
《7》	Health and medicine	2 words

The number of loanwords taken into account is small (34), and they may not by necessarily classified into these seven separate groups as listed above. These semantic

[12] Wikipedia "Wheat gluten"/グルテンミート. Available at https://ja.wikipedia.org/wiki/グルテンミート (accessed 17. 9. 2018).

classes were named just according to my personal intuition, though existing namings of concepts were also taken into account.[13] The classes 《2》 and 《3》 for example, are conceptually quite close, but the concepts of rather old origin (up to the end of Edo period) are grouped together under 《3》, and more or less "neutral" namings of sports and martial arts are in 《2》. The semantic class of "Art" 《4》 is also very various, from the traditional Japanese art of *ikebana* to new popular culture of *anime* and *manga*, as well as from such visual art to literal art of *haiku*, and so on. It is also interesting that the board game *go* is popular among Slovene people since 1960s[14], but next to this rather traditional game, there are new games of *karaoke* and *sudoku*, of which the namings are already accepted into Slovene.

In relation to the semantics of the loanwords listed in Table 2, special attention shoud be paid to the dictionary entry of the word *cunami.* The entry lists two separate meanings, one for the physical phenomenon of tsunami, the second for an expressive use of this word in context (the following example in jounalism): *žal se nam dogaja cunami odpuščanja delavcev, zlasti delavk* (="a tsunami of dismissing workers, especially female workers, is unfortunately happening to us"). Such expressiveness of the users of loanwords is a vital reason for borrowed words to be stabilized as a lexical entry in the recipient language in the process of borrowing.

4 Image of Japan through loanwords

Image of Japan among Slovenes may be given by the types of loanwords in Slovene. Since the phonetic and orthographic appearance of loanwords are adjusted to the environment in the Slovene context, their "Japaneseness" do not always persist. Particularly in case of 'happily' slovenized words, some speakers and users of Japanese may be surprised to know their origin.

For the creative competence of the local people borrowing words of foreign origin, the following may be confirmed[15]:

1. For the effective use of loanwords in communication, grammatical knowledge of the source language is not required.
2. Foreign words are borrowed into the existing grammatical framework of the recipient language. Users of the recipient language find a proper place for each new word, if and only if they have a semantic reason for borrowing.

The close analysis of dictionary entries revealed some interesting cases of integration of Japanese concepts, customs, cultural and natural phenomena into the

[13] For example, the list of Slovene words of Japanese origin mentioned under footnote 9 above.

[14] http://www.go-zveza.si/gzs/go-drustvo-ljubljana (Society of Go in Ljubljana, accessed, 16. 9. 2018)

[15] Also in Shigemori Bučar (2011, pp. 259-260).

language practice in Slovenia. The practice of borrowing from Japanese into Slovene does not have a very long history. But we discovered some older cases of borrowing, though with the help of other European languages, e.g. *soja* going back to 17th century (?), *džiudžicu* certainly going back to the time before 1946. Image of Japan, at least through the loanwords, is in the area of culinary experience and creation (7 headwords classified under《1》Cooking and food) and in traditional sports which are surprisingly popular and persistent among Slovenes (6 headwords under《2》Sports and martial arts). Other semantic classes showed mixed nature of old and new, haiku and anime, kamikaze and tsunami, go and sudoku.

In this rapidly changing world, it would be exciting to see more derivations with originally Japanese elements (*sumoborec*, *kamikazni kombi* = "suicidal minivan") and metaphorical and expressive use of loanwords ("a tsunami of dismissing workers").

References

FRAN (2017). Dictionaries of the Fran Ramovš Institute of the Slovenian Language ZRC SAZU Version 5.0. Ljubljana: ZRC SAZU. Available at https://fran.si/

Golob, N., & Petrovčič, M. (2018). Hokkaido Pumpkins and Huawei Phones: Anti-hiatus Tendencies in Slovene. *Acta Linguistica Asiatica*, *8*(2), 63-82. https://doi.org/10.4312/ala.8.2.63-82

Mlakar, B. (2009). Pregled sistemov latiničnega zapisa japonskega jezika, Predlogi za zapisovanje in pregibanje besed iz japonščine in kitajščine. *Azijske študije, 13*(2), 26-38.

Priestly, T. M. S. (1993). Slovene. In B. Comrie & G. Corbett (Eds.) *The Slavonic languages* (pp. 388-451). London: Routledge.

Shibatani, M. (1990). The Japanese Language 8: Phonology. *The Languages of Japan* (pp. 158-184). Cambridge: Cambridge University Press.

Shigemori Bučar, C. (Ed.) (2009). *Predlogi za zapisovanje in pregibanje besed iz japonščine in kitajščine*. Ljubljana: Oddelek za azijske in afriške študije, Filozofska fakulteta, Univerza v Ljubljani.

Shigemori Bučar, C. (2011). Creative competence in borrowings: words of Japanese origin in Slovene. *Linguistica*, *51*, 245-262.

Shigemori Bučar, C. (2012). Gendai surovenia-go ni okeru nihongo kara no gairai-go — keitai-teki kategorī to washa no sōzō-sei [現代スロヴェニア語における日本語からの外来語—形態的カテゴリーと話者の創造性]. *Proceedings of the 2nd International Symposium of the Department of Asian Studies, Faculty of Arts, University of Ljubljana* (pp. 36-41). Ljubljana: University Press.

Toporišič, J. (2000). *Slovenska slovnica.* Maribor: Založba obzorja.

Abbreviations

[dat]	dative case
Eng.	English
Etym.	Slovenian Etymological Dictionary
f	female
Fr.	French
[gen]	genitive case
Germ.	German
[inst]	instrumental case
It.	Italian
m	male
[nom]	nominative case
mn.	plural
SNB	Dictionary of New Slovenian Words
SSKJ	Dictionary of the Slovenian Standard Language

UDERSTANDING SARCASTIC METAPHORICAL EXPRESSIONS IN HINDI THROUGH CONCEPTUAL INTEGRATION THEORY

Sandeep Kumar SHARMA
Indian Institute of Technology Patna, India
skvpsharma@gmail.com

Sweta SINHA
Indian Institute of Technology Patna, India
apna1982@gmail.com

Abstract

Metaphorical expressions are one of the most indispensable aspects of human language, thought and action. Their meanings are figurative, which in other words means that they carry literal meanings that are in direct opposition to the intended or primary meanings. The usage of metaphors is not limited to figurative writing and speaking only but they are pervasively found in everyday language. Irony, sarcasm, jokes, puns and other such metaphorical expressions rampantly occur in our everyday speech. This paper examines the abstract notion of sarcasm within the framework of conceptual integration theory, and with special reference to Hindi language. A corpus of five thousand sentences has been procured from Indian Language Technology Proliferation and Deployment Centre (TDIL) for the present study. The findings aim to provide a theoretical understanding of how Hindi sarcasm is perceived among the native speakers.

Keywords: cognitive linguistics; metaphor; sarcasm; irony; conceptual blending; figurative language

Povzetek

Metaforični izrazi so eden nepogrešljivih vidikov človeškega jezika, mišljenja in delovanja. Njihovi pomeni so preneseni in so lahko v popolnem nasprotju z izvornimi pomeni pozameznih besed v izrazih. Metafore niso sredstvo samo v literarnem jeziku, ampak so splošno razširjene tudi v vsakodnevnem pisnem in govornem izražanju. Ironija, sarkazem, šale, besedne igre in drugi metaforični izrazi so vsakodnevnica v govoru. Članek preučuje abstaktnost sarkazma na primeru hindujščine in sicer po teoriji konceptualne integracije. Za raziskavo je bil uporabljen korpus Indijskega centra za jezikovne tehnologije (TDIL) s pet tisoč primeri stavkov. Rezultati raziskave predstavljajo teoretično razumevanje sarkazma, ki ga uporabljajo hindujski govorci.

Ključne besede: kognitivna lingvistika; metafora; sarkazem; ironija; pojmovno prekrivanje; figurativni jezik

Acta Linguistica Asiatica, 9(1), 2019.
ISSN: 2232-3317, http://revije.ff.uni-lj.si/ala/
DOI: 10.4312/ala.9.1.89-103

1 Introduction

The field of cognitive linguistics has generated a powerful set of theoretical tools for analyzing the ways in which we understand, communicate and create concepts. Development of the conceptual theory has brought an insight into the appearance and usage of metaphorical expressions in everyday speech. Conceptualization of a metaphor is grounded in wide range of bodily, social and cultural experiences, and creates an integral aspect of cognitive faculty which plays a creative role in meaning construction of knowledge and understanding reality. A metaphorical expression is one of the most indispensable aspects of human life including language, thought and action. As Lakoff and Johnson (1980) noted "metaphor pervades our normal conceptual system. Because so many of the concepts that are important to us are either abstract or not clearly delineated in our experience (the emotions, ideas, time etc.), we need to get a grasp on them by means of other concepts that we understand in clearer terms" (Lakoff & Johnson, 1980, p. 115). Therefore, Lakoff and Johnson argue that "metaphor is a natural phenomenon" (Lakoff & Johnson, 1980, p. 247), it is beyond language as it is found primarily in thought and action (Lakoff & Johnson, 1980, p. 153). It reflects a particular speech community and its creative aspects of language and culture in a positive as well as negative light. Metaphoricity is a specific feature of human language where no form of language can exist without metaphorical traits (Goalty, 1997).

Cognitive linguists claim that metaphors are not only limited to figurative writing. They are thought to be a specific mental mapping that reflects how people think and imagine in everyday life (Lakoff & Johnson, 1999). Irony, satire, sarcasm and other such metaphorical expressions rampantly occur in everyday conversation of different speech communities (Tay, 2014).

Sarcasm seems to stand out due to its heavily negative intention (Joshi, Bhattacharyya, & Carman, 2017). It is thought to be a form of figurative language and an integral part of human discourse where literal meaning of words are in direct opposition to the intended meaning (Grice, Cole, & Morgan, 1975). Under the developmental approach, sarcasm is described as culturally salient phenomenon that offers a clear cut case of discrepancy between content and literal meaning (Prokofiev, 2017). Sarcasm uses wit, ridicule and mockery. It is a form of a metaphorical expression which is identified by literary scholars as a skill of using incongruity to indicate distinction between reality and expectation. Sarcasm is not to be confused with irony, which pertains to situation and is thought be a tool for expressing sarcasm. Sarcastic language is defined as 'irony that is especially bitter and caustic' (Gibbs, 1994). Discrepancies between irony and sarcasm are reported to include disparity of literal meaning of an utterance – positive or negative, where a positive literal meaning is subverted by a negative intended meaning (Dews & Winner, 1995). In this respect, to understand sarcasm it would be crucial to understand the information that violates the truthfulness (Gibbs Jr & O'Brien, 1991). In the light of differential description between

sarcasm and irony, when one comes to the conceptualization of sarcasm and its function, there are several paradigms that function as a conceptual framework to understand mental representation of metaphorical expressions in the process of meaning construction.

Basically it seems that any metaphorical expression can be analyzed through Lakoff's theory of conceptual metaphor, where metaphorical statements are largely perceived through one-to-one mapping. However, not all such statements can be perceived through the conceptual metaphor theory alone because their lexical extensions go beyond what appears on the surface level. Therefore, for the conceptualization of extended meaning, we should focus on the conceptual framework of integration or blending theory, which regulates the process of conceptualization in human cognition in the form of novelty construction as well as the understanding of one idea or conceptual domain in terms of another. In conceptual integration or blending theory, knowledge structures or mental spaces are selectively projected into blending space in which projected conceptual elements are assimilated to create a novel concept with respect to content based emergent structure. Constructing a new meaning through the integration of existing concepts provides a wide range of conceptual concepts of metaphorical expression from two input spaces into a new mental extent called blended space. This is the creativity of cognitive enterprise that frequently displays an emergent structure of conceptual relations that are unavailable in input spaces (Fauconnier & Turner, 2002).

This approach is an effective mental process of composition, completion and elaboration of the blend. At the cognitive stage, a conceptually integrated emergent structure comprises neuro-biologically based semantic meanings with generative grammar, which relates them. In the theory of conceptual integration, it is not not a word, sentence or objects but rather the meaning that evokes an effective mental process. The following figure has been taken from Fauconnier and Turner's "basic diagram" to represent the cognitive operation of conceptual blending (Fauconnier & Turner, 1998). It is a visual illustration of cognitive process and the construction of a concept, emerging from blending two input spaces into a single blended space. A special reference to Hindi language is added.

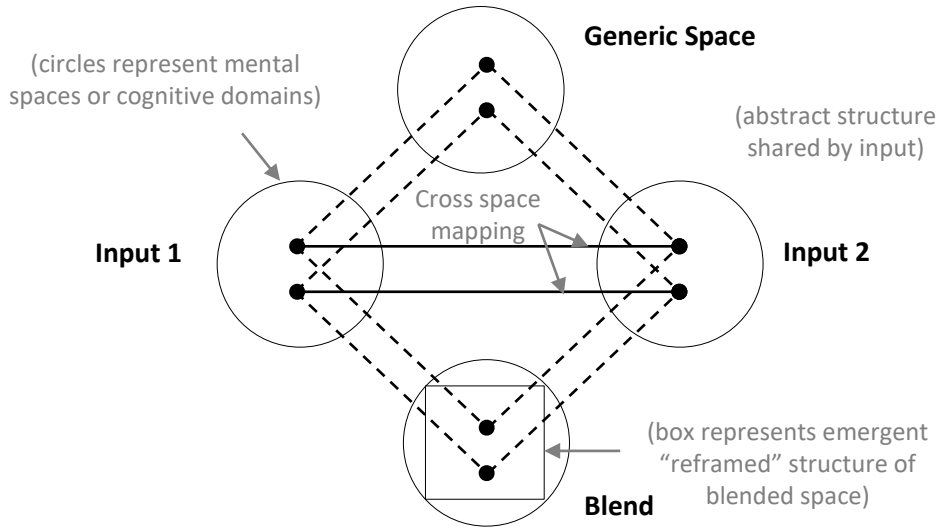


Figure 1: Conceptual integration "basic diagram" (Fauconnier & Turner, 1998)

2 Sarcasm as a "linguistic element"

A landmark article on psycholinguistics of sarcasm by Gibbs (1994) begins with the definition of irony from the Oxford English dictionary as "the use of words to express something other than and especially the opposite of literal meaning". Both irony and sarcasm serve multiple purposes in spoken language, especially in discourse. Speakers and writers have been found to use sarcasm to convey humor and censure, and to indicate social norms violation (Brown & Levinson, 1978; Kreuz & Glucksberg, 1989; Shamay-Tsoory, Tomer, & Aharon-Peretz, 2005). Sarcasm often employs contradiction between literal and intended meaning, and appears as a universal feature of human language. Different speech communities use different visual, prosodic, lexical and contextual cues to convey sarcastic messages in distinct ways and with different emotions, attitudes, or intensions. Moreover, the function of sarcasm varies across speech communities consequently identifying and appropriately interpreting sarcastic utterances in a particular language different from others. Doing so entails both, the ability to read an array of potential cues that mark an utterance as sarcastic as well as the capacity to correctly interpret a speaker's motivation for using sarcasm (i.e. ridicule or insult).

Sarcasm is ubiqutous in all natural languages including Hindi and the speakers seemingly use various verbal and non-verbal strategies to express themselves. Non-verbal behaviors are also used to signal sarcasm such as raised eyebrows, eye rolling and titled head. The saliency of sarcastic indicators depends on how explicitly a speaker wishes to mark sarcastic utterances as a speaker may intentionally restrict the number

of cues offered in order to obfuscate subversive motives and preserve deniability. Moreover, contextual cues including biographical information, physical setting and even the history of the relationship between interlocutors can figure into sarcasm use. Gibbs and O'Brien (1991) describe that the violation of truthfulness maxim is the key to understanding sarcasm. The intended meaning of sarcasm can not be understood until the listener observes literal meanings of the text that violates truthfulness. Grice (1975) points out the exploitation of maxim which is observed by means of metaphorical expressions. Clark and Haviland (1977) claim that a deliberate violation is perceived by the hearer while interpreting what the speaker intended to say. Thus, sarcasm prevails through various dimensions such as a failed prediction, insincerity of pragmatic context, negative intention etc. (Campbell & Katz, 2012).

3 Method of Analysis

Based on the theoretical description provided in the sections above, the methodology section outlines the research process from planning to presentation through qualitative approach. Data collection used is an annotated digital corpus[1] that has been procured from ongoing and completed projects to strengthen technology development in and for Indian languages. For the present study five thousand sentences have been extracted from the large chunk of corpus which was built to investigate sarcasm in native speakers through conceptual integration theory. The data have been procured from the Indian Language Technology Proliferation and Deployment Centre.[2]

The data belongs to the discourse domain of politics that has already been segregated in the corpus. To filter the required data set, five thousand sentences have been read manually by a native speaker of Hindi. The filtered data sets have been analyzed qualitatively within the framework of conceptual integration to analyze sarcastic expressions. For the understanding of the context, the data sets have been analyzed through Grice's maxim of conversation that examines the nature of congruency and incongruence (Grice, 1975). The flow chart below (Figure 2) gives a schematic representation of the methodology adopted for this study.

[1] Indian Language Technology Proliferation and Development Centre, TDIL (Technology Development for Indian Languages). www.tdil.meity.gov.in

[2] Ministry of Electronics and Information Technology - MeitY. The Centre works for consolidating and making available the linguistics resources under the initiation of Technology Development in Indian Languages Programme of MeitY.

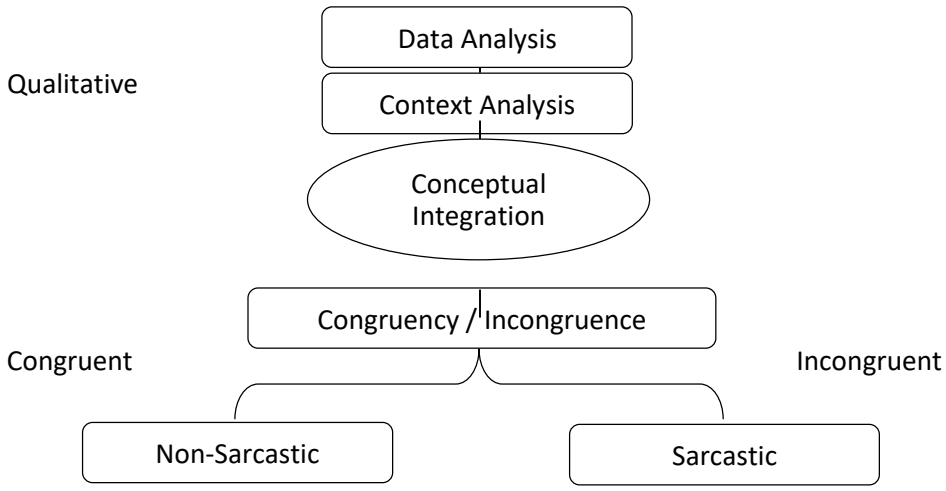


Figure 2: Methodology for investigating sarcasm in Hindi though Conceptual Integration

The current research consists of six sections. The first introductory section attempted to introduce the research area by providing a background of similar researches that have already been done. The section also highlighted the theory of conceptual blending/integration which has been found to be very relevant in the conceptualization of figurative language in cognitive linguistics. Focusing this research on Hindi sarcasm and its conceptualization, section 2 attempted to establish sarcasm as a linguistic element. The current section (section 3) has already outlined the methodology/tools adopted for the purpose of analysis. The following section (section 4) analyses 12 sarcastic Hindi sentences that have been procured from the data source. Analyses as in examples (1) to (12) indicate two inputs spaces, that is input 1 and input 2 as being juxtaposed and creating a blended space of incongruence entities. This incongruity yields sarcasm. Section 5 discusses the examples in accordance with the theory of conceptual blending or integration and the results are eventually summarized in section number 6.

4 Analysis and results

Extracted sarcastic sentences (1) to (12) have been analyzed through conceptual blending to observed the integrating process of a novel concept. The two input spaces contain the conceptual element of a particular metaphor. Both these input spaces have developed across their space mapping relations in order to obtain the correct perception of their conceptual constituent. As these constituents blend in a space, a naturally incongruous disparity is observed.

(1) chhata na hua Chhatrapati ka chhatra ho gaya
umbrella NEG be Chhatrapati POST.P umbrella be MASC. PST PVF
'As if this is not an ordinary umbrella but the umbrella of Chhatrapati Shivaji.'

Context: The expression is perceived as sarcasm because an ordinary object is attributed an extraordinary/unique status.

Input 1 Chhatrapati ka chhatra	**Input 2** Chhata	**Generic space**	**Blend**
Historical identity	Non- historical identity	An object	Historical – Non historical
Superior	Public approach/General	Concept	Kingship – Public
Royal use/Extraordinary	Ordinary use	Purpose	Royal – Ordinary
Symbol of prestige	Symbol of need	Symbol	Prestige – Need

Sarcastic Blend: Ordinary things cannot be a royal icon.

(2) aira-gaira nahi kale angrejon ka sartaaj aa raha tha
stranger NEG black Englishmen POST.P sartaaj come MASC. PST. PROG
'The one who is coming is not a nobody but the king of black Englishmen.'

Context: This is perceived as sarcasm in a situation when a brown skinned person behaves like a ruler amongst their own people.

Input 1 Aira-Gaira	**Input 2** Kale angrejon ka sartaaj	**Generic space**	**Blend**
Undistinguished identity	Distinguished identity	Agent	Undistinguished – Distinguished
Irrelevant attitude	Relevant attitude	Position	Irrelevant – Relevant
Inglorious position	Glorious position	Knowledge	Inglorious – Glorious
Intellectual instability	Intellectual stability	Behavior	Instability – Stability

Sarcastic Blend: One who behaves like a Britisher. (Britain had colonized India for around 200 years.)

(3) protocol me adami adami nahi rahta kenchua ban jata hai
protocol POST.P man man NEG live earthworm become.MASC.PRS
'A human behaves more like an earthworm when following protocols.'

Context: A sarcasm on the government system which reduces work efficiency by sticking to protocols too strictly.

Input 1 Adami	**Input 2** Kenchua	**Generic space**	**Blend**
+Human	–Human	Agent	+Human – –Human
Mobility	Slow mobility (creeping)	Movement	Mobility – Slow mobility

Sarcastic Blend: Protocols reduce work efficiency of a man.

(4) lalu ji to media ke darling hai
Lalu HON CONJ media POST.P darling be PRS.
'Lalu Ji is loved by the media.'

Context: Despite so many events of national importance, the media has maximum coverage of Lalu Yadav.

Input 1 Lalu	**Input 2** Darling	**Generic space**	**Blend**
Public figure	Personal image	Agent	Public – Personal
Political power	Non-political power	Power	Political – Non political
Social representative	Individual supporter	Favor	Social – Individual

Sarcastic Blend: Getting publicity without reason.

(5) Rahul Gandhi ko yuva neta ghoshit karne matra se kya yuva kangres ko wot denge
Rahul Gandhi POST.P youth leader announcedo only POST.P what youth congress POST.P vote give MASC. FUT
'Will youth vote for Congress only by announcing Rahul Gandhi as a youth leader.'

Context: For vote bank of youth, Congress Party announced Rahul Gandhi as a youth leader who has crossed the age of youth.

Input 1 Rahul Gandhi	**Input 2** Yuva	**Generic space**	**Blend**
By age 47 yrs.	By Age 15-29 yrs.	Biological aspect	Age: 47 – 15-29 yrs.

Sarcastic Blend: To be called youth one must possess the quality of youth.

(6) bhajpa ne Narendra Modi ko sankatmochan bana ke bheja hai
bhajpa NOM narendra modi ACC troubleredeemer make POST.P send MASC.PRS
'Narendra Modi is sent as trouble redeemer by Bhajpa.'

Context: This is expressed sarcastically because Bhajpa represents Narendra Modi as the Lord Hanuman who solves the problems of people as per Hindu Mythology.

Input 1 Narendra Modi	**Input 2** Sankat Mochan	**Generic space**	**Blend**
Charismatic leader	Charismatic lord	Ability	Leader – Lord
Reformation of country	Savior of universe	Conscientious	Reformation - Savior
Circumscribed	Omnipresent	Presence	Circumscribed - Omnipresent
Positional act	Ubiquitously act	Power	Positional- Ubiquitously

Sarcastic Blend: Problems of the party are too huge for Narendra Modi to tackle.

(7) tejashwi jaise neta Baisakhi ke sahare rajniti me aate hai
tejashwi ADV leader crutches POST.P support politics POST.P come MASC.PRS
'Leaders like Tejashwi come into politics through crutches.'

Context: Just as crutches enhance the mobility of those who have certain physical incompetence, similarly family has supported Tejashwi to join politics.

Input 1 Tejashwi	**Input 2** Baisakhi	**Generic space**	**Blend**
By support	For support	Sustain	Support: By – for
Public strength	Strength of helpless	Power	Strength: Public – Helpless
Political upliftment	Miserable upliftment	Goal	Upliftment: Political – Miserable
Legacy Support	Moral support	Need	–

Sarcastic Blend: One who rises in politics through support.

(8) Tejashwi pahle padhai kare fir mange hisaab.
tejaswi before study do CONJ ask justification
'Let Tejashwi study first then ask for justification.'

Context: Tejashwi talks without knowing the context.

Input 1 Tejashwi	**Input 2** Hisab	**Generic space**	**Blend**
Surface understanding	Deeper understanding	Comprehension	Understanding: Surface – Deeper
Political discourse	Educational discourse	Communication	Educational – Political
Discrete knowledge	Integrate knowledge	Intelligence	Knowledge: Discrete – Integrate

Sarcastic Blend: Without education intellectual skills cannot be improved.

(9) Daru bina dosti nahi tikti.
liquor without friendship-NEG sustain FEM.PRS.IMP
'Without liquor friendship does not sustain.'

Context: Liquor is important in friendship in the sense that friendship can be sustained longer.

Input 1 Daru	**Input 2** Dosti	**Generic space**	**Blend**
Loss of reasoning	Help in decision	Catalyst	Loss – Help
Unhealthy habit	Healthy Behavior	Effect	Healthy – Unhealthy
Short term pleasure	Timeless companion	Bond	Short term – Timeless
Loss of confidence	Boost confidence	Action	Loss – Boost

Sarcastic Blend: For friendship to continue one needs to offer liquor to friends.

(10) Hum sab to Hindustan me fail hue lekin tumne to Londonme
Hum sab to Hindustan me fail hue lekin tumne ADV landanPOST.P
fail ho-kar dikha diya.
fail be-CP see give MASC.PST
'We failed in Hindustan but you showed up having failed in London.'

Context: It is perceived as sarcasm when someone who had failed in their native land due to lack of facilities compares themselves to those who studied abroad but could not pass examination despite having all facilities at hand.

Input 1 Hindustan	**Input 2** London	**Generic space**	**Blend**
Less infrastructure	Great infrastructure	Facility	Infrastructure: Less – Great
Non practical implication	Practical implication	Uses	Non practical – Practical
Trend based admission	Interest based admission	Selection	Trend based – Interest based
Believe in grade	Believe in skill	Understanding	Believe: Grade –Skill

Sarcastic Blend: In spite of all the facilities he could not pass.

(11) Dimagme kitabe bharne sejebe nahi bharti hai.
Brain POST.P book fill POST.P.pocket.PL NEG fill FEM.PRS
'Just reading the books does not make one rich.'

Context: Only reading is not enough to earn money. One also needs to perform.

Input 1 Dimag me kitabe bharnaa	**Input 2** Jebe bharnaa	**Generic space**	**Blend**
Research skill	Earning skill	Ability	Skill: Research – Earning
Intellectual approach	Realistic approach	Perception	Approach: Intellectual –Realistic
Use knowledge	Capitalize passion	Decision	Use Knowledge – Capitalize passion
Creative thinking	Business mind	Comprehension	Creative thinking – Business mind

Sarcastic Blend: To make money, work has to be done.

(12) Ab yogi bhi bhogi ki tarah bina khaye nahi rah sakte.
now saint ADV bhogi like without eat MASC.PST . NEG live can.
'Now ascetics too used to eat like common men to survive.'

Context: The line of distinction between an ascetic and a common man has blurred with regards to a way of life.

Input 1 Yogi	**Input 2** Bhogi	**Generic space**	**Blend**
Devine love	Worldly love	Feeling	Devine – Worldly
Undesirable	Desirable	Need	Undesirable – Desirable
Inner happiness	Physical happiness	Satisfaction	Happiness: Inner – Physical
With equanimity	Without equanimity	Balance	Equanimity: With – Without

Sarcastic Blend: Nowaday's Yogi became Bhogi.

5 Discussion

Above examples were obtained from an annotated digital corpus of five thousand sentences, which have been read to extract the sarcasm-oriented utterances from the large chunk of annotated data set.

The domain mapping of conceptual metaphor does not always recognize all metaphorical expressions as it focuses only on one-to-one mapping of source and target domain. Therefore, to grasp extended meanings of metaphorical expressions it was important to go through either conceptual blending or integration theory to achieve an integrated mechanism with which observation of the novel construction gets possible.

Input spaces contain one or more conceptual elements of a particular metaphor and represent its attached construal aspects. These conceptual packets have been observed with reference to generic sense. It was found out that they may have a general or abstract structure, which are seemingly shared by both input spaces to express a common sense for different conceptual constituents. Input space may project into blended space, and as such represent an emergent structure of a novel concept.

As in (9) the two input spaces *Daru (liquor)* and *dosti (friendship)* have different conceptual elements where they have an abstract generic sense of *catalyst, effect, bond, and action* with respect to each conceptual element respectively. The cross space mapping between input one (*loss of reasoning, unhealthy habit, short term pleasure and loss of confidence*) and input two (*helps in decision, healthy behavior, timeless companion and boost confidence*) project their conceptual elements into the blended space that created an emergent structure of conceptual meaning of metaphorical expressions in blended space. Out of 5,000 sentences procured from data source, 12 sentences needed to be explained through conceptual blending or integration theory.

With respect to the above description it is observed that negative sentences occur not only in the intension of an individual but also reflect literally, and in the form of dropped negation in sarcastic expression, as a cue. Such indicatory cues help develop the understanding of sarcastic expression. Besides such sarcastic expressions there are affirmative, interrogative and imperative sentences also that are used as tools to ridicule an individual.

As such, sarcasm can be described as an obscure phenomenon. It carries several functions and uses means that are different from other communicative acts. The functional approach used in this research enables one to observe the intension of sarcastic utterances.

Sarcasms in speech are used to express intense emotions. Based on the above results it is observed that sarcasm can be used in both positive and negative sense. The positive use of sarcastic utterance is attached with humorous intent through which an individual makes a critical comment without appearing rude. Sentences (1), (4) and (10) are positively functional. Sarcasms with a negative function may be realized in various ways such as through ridicule, indirect rebuke, minor irritation etc. They are used to makes critical remarks. In this respect sentences (2), (3), (5), (6), (7), (8), (9), (11) and (12) are negatively functional.

In short, sarcasm has several functions and they rampantly occur in the discourse of people on everyday basis, primarily – though not necessarily – with negative function.

6 Conclusion

Mental processes have always intrigued mankind. Despite numerous researches that have already been conducted, a large part of this area is still to get examined carefully.

Cognitive linguistics is an upcoming area in linguistics. The field of cognitive linguistics has generated a powerful set of theoretical tools for analyzing the way in which we understand, communicate and create concepts. The development of the conceptual theory has brought the ubiquity of metaphorical expressions in everyday speech.

The conceptualization of a metaphor is grounded on a wide range of bodily, social and cultural experiences that create an integral aspect of cognitive faculty and play a creative role in meaning construction as well as in understanding reality. Metaphorical language is an indispensable aspect of human life through which people use figurative language to represent abstract concepts with reference to concrete entities for easy comprehension. In this respect, sarcasm is a form of figurative language and integral part of human discourse where literal meaning of words are in direct opposition to the intended meaning, which is pervasively used in everyday language to ridicule someone.

Hindi sarcastic statements are heavily dependent on the contextual knowledge of the hearer in order to be effective. Conceptualization of sarcastic metaphorical expressions in Hindi can not be explained by conceptual metaphor theory alone. Major sarcasms – or to say more demanding ones – are those that can only be analyzed through blending/integration, which consequently brings one to the conclusion that not all metaphorical expressions can be understood through the help of conceptual theory or one-to-one mapping relationship only. To get the extended meaning we need to look at the theory of conceptual integration or blending.

This paper tried to investigate the conceptualization of sarcastic expressions in Hindi language within the framework of conceptual integration. Conceptual blending integrates the conceptual elements into blended space with the help of generic sense and gives an emergent structure of meaning to observe its functions and uses. Hindi sarcasms, like figurative linguistic tools of other languages, are deeply rooted in the historical and cultural evolution of the language and its speakers. An effective comprehension and conceptualization of such concepts needs a multi-layered cognitive theoretical approach similar to what has been studied in this paper.

References

Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction* (pp. 56-311). Cambridge: Cambridge University Press.

Camp, E. (2012). Sarcasm, pretense, and the semantics/pragmatics distinction. *Noûs, 46*(4), 587-634.

Campbell, J. D., & Katz, A. N. (2012). Are there necessary conditions for inducing a sense of sarcastic irony? *Discourse Processes, 49*(6), 459-480.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Erlbaum.

Dews, S., & Winner, E. (1995). Muting the meaning A social function of Irony. *Metaphor and Symbol, 10*(1), 3-19.

Eisterhold, J., Attardo, S., & Boxer, D. (2006). Reactions to irony in discourse: Evidence for the least disruption principle. *Journal of Pragmatics, 38*(8), 1239-1256.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive science, 22*(2), 133-187.

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities.* New York: Basic Books.

Gibbs Jr, R. W., & O'Brien, J. (1991). Psychological aspects of irony understanding. *Journal of pragmatics, 16*(6), 523-530.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding.* Cambridge: Cambridge University Press.

Giora, R. (1995). On irony and negation. *Discourse processes, 19*(2), 239-264.

Goalty, A. (1997). *The language of metaphor.* London/New York: Routledge.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and Semantics* (Vol. 3, pp. 41-58). New York: Academic Press.

Grice, H. P., Cole, P., & Morgan, J. L. (1975). *Syntax and Semantics.* New York: Academic Press.

Joshi, A., Bhattacharyya, P., & Carman, M. J. (2017). Automatic sarcasm detection: A survey. *ACM Computing Surveys (CSUR), 50*(5), 73.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of experimental psychology: General, 118*(4), 374.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202-251). New York, NY, US: Cambridge University Press.

Lakoff, G. J., & Johnson, M. (1980). *Metaphors We Live By.* Chicago/London: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought* (Vol. 28). New York: Basic Books.

Prokofiev, G. (2017). Differentiation between irony and sarcasm in contemporary linguistic studies. Вісник Дніпропетровського університету імені Альфреда Нобеля. *Серія: Філологічні науки, 13*(1), 233-239.

Ramos, F. Y. (2000). Literal/non literal and the processing of verbal irony. *Pragmalingüística* (8-9), 349-374.

Shamay-Tsoory, S. G., Tomer, R., & Aharon-Peretz, J. (2005). The neuroanatomical basis of understanding sarcasm and its relationship to social cognition. *Neuropsychology, 19*(3), 288.

Tay, D. (2014). Lakoff and the Theory of Conceptual Metaphor. In J. Littlemore, & J. R. Taylor (Eds.), *The Bloomsbury Companion to Cognitive Linguistics* (pp. 49-59). London, UK: Bloomsbury.

AFFECTION OF THE PART OF SPEECH ELEMENTS IN VIETNAMESE TEXT READABILITY

Điệp Thi Nhu NGUYỄN
Vietnam National University Ho Chi Minh City, Vietnam
nhudiep2004@gmail.com

An-Vinh LƯƠNG
Vietnam National University Ho Chi Minh City, Vietnam
anvinhluong@gmail.com

Điền ĐINH
Vietnam National University Ho Chi Minh City, Vietnam
ddien@fit.hcmus.edu.vn

Abstract

While English text readability has been studied for a long time, investigating text readability in Vietnamese, a low-resourced language with poorresearch technologies and data sets questionable of international importance, is at its beginnings. In readability research, it is generally the "word" that has been carefully investigated. Based on the comparison of elements affecting readability of the "word" unit in English, we determine the parts of speech (POS) in Vietnamese that were found to influence Vietnamese text readability. In this study, prose texts in Vietnamese textbooks at different difficulty level were taken as the data to find out the POS frequencies and their correlations. In terms of frequency, our findings can initially assist users when editing documents, reforming textbooks, and question banks for native Vietnamese in general and foreigners in particular. Even more important, with these findings we can identify those linguistic elements that are considered the "potential" POS affecting Vietnamese text readability, and make grounds for further studies.

Keywords: text readability; parts of speech; Vietnamese textbooks; elementary level

Povzetek

Medtem ko je že precej vemo o bralni pismenosti angleških tekstov, pa so takšne raziskave na tekstih v vietnamščini šele na začetku. Večina raziskav o bralni pismenosti se osredotoča na "besedo". Na osnovi primerjav elementov, ki vplivajo na bralno pismenost na nivoju besede v angleščini, smo v naši raziskavi določili besedne vrste

Acta Linguistica Asiatica, 9(1), 2019.
ISSN: 2232-3317, http://revije.ff.uni-lj.si/ala/
DOI: 10.4312/ala.9.1.105-118

(angl. “parts of speech”, POS), pri katerih smo zaznali, da vplivajo na bralno pismenost v vietnamščini. V raziskavi so bili obravnavani učbeniki vietnamščine in sicer njihovi prozni teksti, iz katerih smo ocenili pojavnost posameznih besednih vrst in njihovo korelacijo z različnimi težavnostnimi nivoji. Že same informacije o pojavnosti lahko pripomorejo k boljšemu razumevanju bralne pismenosti in so v pomoč pri pripravi in urejanju dokumentov, pisanju učbenikov, sestavljanju vprašalnikov tako za domače govorce, še posebej pa za tuje govorce vietnamščine. Še bolj pomembni pa so seveda pridobljeni podatki o jezikovnih elementih, ki so označeni kot besedne vrste, ki potencialno vplivajo na bralno pismenost v vietnamščini. Slednji predstavljajo osnovo za vse nadaljne raziskave.

Ključne besede: bralna pismenost; besedne vrste; učbeniki vietnamščine; začetna stopnja

1 Introduction

The studies of readability have been done since the early nineteenth century. Among these achievements are the formulas for measuring readability, which are used as a tool for determining the complexity of the text. Therefore, they can help users select an appropriate text with different reading levels for the readers in efficiently, saving time and labor. The results of the research have applied in various areas of society, such as the integratedly measuring the Flesch formula in Microsoft Office software or the same with the formulae: Flesch-Kincaid, Cohmetrix, Idicies, Lexile Measures, etc. in the Common European Framework of Reference.

In forming a formula or a tool to measure text readability, linguistic elements or linguistic components in a particular text play a very important role, as shown in a lot of readability research, such as Gray and Leary (1935), Lorge (1939), Rudolf Flesch (1943; 1946; 1948), Graesser et al. (2004), and McNamara et al. (2014). These linguistic elements were gained through analyses on the shallow/surface features on one hand, such as the average length of words by the number of syllables, the average numbers in a sentence, or the frequency of words; and the deep features of the language on the other hand, such as the parsed syntactic features, the language modeling features, or the part of speech- based features.

With the scope of this article, we first define the part of speech as the linguistic elements affecting the text readability in Vietnamese based on the contrast of the linguistic elements affecting text readability in English, we survey and evaluate readability influences of part of speech (POS) elements of prose texts in Vietnamese subject textbooks for elementary school-aged children based on the several statistic measures.

Results of this study are expected to be useful to writers, editors, and especially to teachers and learners of Vietnamese, who compile or select lectures and banks of questions based on the grade level.

2 Methodology and corpus

Our corpus represents prose texts in Vietnamese textbooks for elementary school children (grades 2–5) that were published by Education Publisher in 2016. In the preprocessing, we have decided to leave out the texts that were in forms of questions, puzzles or drawing annotations, and therefore were left with 209 texts in the end. Those texts are all estimated to provide children with general knowledge and help them practice reading skills. Linguistic elements with surface features are described in Table 1 below:

Table 1: Vietnamese textbook corpus

Grade	Number of Texts	Number of Words	Number of Sentences
2	67	57 – 251	5 - 40
3	62	112 - 279	8 - 35
4	40	144 - 520	7 - 47
5	40	111 - 381	4 - 52

We used the "CLC-Vietnamese-Toolkit"[1], generated by Computational Linguistics Center, University of Science, HCMC, to handle the POS in each text, and calculate their frequency. Besides, the relationship between the POS with the text readability was also investigated.

3 Affection of linguistic elements in text readability

3.1 Linguistic elements affecting text readability in English

Gray and Leary's (1935) identified 288 elements affecting English text readability, and these elements were classified into four main categories: (I) format or mechanical features, (II) general features of organization, (III) style of expression and presentation, and (IV) content (Gray & Leary, 1935).

Within the scope of their study, they have identified 82 language elements that function as the "potential elements" affecting text readability by investigating the linguistic elements of style of expression and presentation alone. These elements are classified under three different units, namely word, sentence and paragraph/passage.

[1] http://www.clc.hcmus.edu.vn/

Among them, 41 elements affecting text readability at word level were counted. With the aim to conduct an experimental research based on quantitative enumeration, 14 out of those 41 language elements were left out of further analysis due to the following reasons: (i) the linguistic elements do not meet the experimental process; (ii) they has not been formed by the clear definitions yet, and (iii) these linguistic elements cannot be measured or counted objectively in largely analyzed cases from the corpus.

Based on this elementary work, many studies have investigated and developed the language elements affecting English text readability. Examining the same "structural elements" as Gray and Leary (1935), Lorge (1939) added an additional variable, "a weighted index of word difficulty". Lorge believed that prepositions played important role to measure syntactic complexity in English. He suggested the readability formula which adjusts weights and uses various combinations of two variables such as (i) prepositional phrases and different hard words, (ii) average sentence length and different hard words, and (iii) the number of prepositional phrases and average sentence length (Lorge, 1939).

In creating a regression formula that could with some accuracy distinguish levels of difficulty for both children's and adults' reading material, besides sentence length Rudolf Flesch (1943) added two other variables: the number of affixes and a variable used in Gray and Leary. The number of personal pronouns, which Flesch limited to gendered (non-neutral) pronouns, were represented by the human interest factor of the texts (Flesch, 1943). Flesch (1948) defined the idea of personal words somewhat differently in order to codify human interest: "All nouns with natural gender; all pronouns except neuter pronouns; and the words people (used with the plural verb) and folks". To this, Flesch added another factor, which he called "personal sentences". This factor was intended to be a measure of the "conversational quality and the story interest" of the passage analyzed (Flesch, 1948). The Art of Readable Writing (Flesch, 1949) was a popular success as a "how-to" book about writing, successful enough that a quarter of a century later the book was reissued in a new, expanded edition (Flesch, 1974). The Reading Ease formula was adapted for use by the United States Military using the same factors but somewhat different weights (Kincaid et al., 1975) and can be found to this day as a tool in the most popular word processing program in the world, Microsoft Word.

Coh-Metrix is a major departure from both the classic formulas and cloze. It is a computational tool that facilitates the formulation and testing of hypotheses about readability and other reading comprehension issues: "Coh-Metrix ... analyzes texts on over 200 measures of cohesion, language, and readability. Its modules use lexicons, part-of-speech classifiers, syntactic parsers, templates, corpora, latent semantic analysis, and other components that are widely used in computational linguistics" (Graesser et al., 2004). In classifying part-of-speech, McNamara et al. (2014) presented that Coh-Metrix permits more sophisticated measures of grammatical complexity, it

can count the mean number of modifiers in noun phrases and the mean number of words that occur before the main verb. In particular, Coh-Metrix includes indices for various linguistic features that can be considered markers of cohesion, for example, it contains an index for measuring the number of causal connectives- connectives indicating the logical relations between parts of the text (e.g., because, so). It also contains an index relating causal particles (e.g., due to, therefore, if) to causal verbs. The hypothesis is that the higher the ratio of causal particles to causal verbs, the more cohesive a text is, since it suggests that there are more explicit indications of how events and actions are interrelated (McNamara, Graesser, McCarthy, & Cai, 2014, pp. 62-68).

Thus, language elements in general, and the parts of speech in particular, have been investigated more and more deeply in English text readability to meet the practical needs. However, it is important to note that there are many differences between English and Vietnamese, ranging from morphological typology (morphemes, word boundaries, the word forms, for example “anh” in Vietnamese means “elder brother” in English), and sentence structure (theme-rheme relationship), to the differences in phonetics and phonology. Therefore, adjustments to the existing model should be made, and comparisons and contrasts between these two languages are crucial in this case (Đinh, 2006). Hence, by comparing the similarities and differences of the linguistic elements between Vietnamese and English in the word unit, this article selects and surveys the POS elements at the word unit from the above-mentioned corpus.

3.2 Linguistic elements affecting text readability in Vietnamese

3.2.1 Lexico - grammatical category

Language vocabularies are generally very large and it is thus reasonable to further divide words into subclasses to make the word-formation rules and those of their usage more comprehensible. There are several ways to do so. For example, words can be further divided in terms of (1) their meanings; namely some words convey one meaning while others are polysemantic, in terms of (2) their origin, where they can be classified into cognates and borrowed words, (3) according to the frequency of their usage, where common, everyday words are used more often than words of slang, dialectal expressions, technical terms, and others. Words can also be divided (4) based on their word-forms into monosyllabic and polysyllabic words, or else into single and compound words, and nonetheless (5) according to their first letter, as in dictionaries.

In Vietnamese, however, there is another crucial way of word classification, which is based on words’ lexical meanings together with their grammatical functions. It is called lexico-grammatical category (Nguyễn, Đoàn, & Nguyễn, 2008, p. 242).

Each grammatical category includes a set of different forms of a word, but each lexico-grammatical category includes a set of words. The process of determining grammatical category generally begins with considering possible forms of a word to determine their number; for example, in English, book (singular) with books (plural). Only then are words categorized into content words (nouns, verbs, adjectives, adverbs…) and function words (articles, prepositions, conjunctions…). On the other hand, applying lexical-grammatical category means that a word carries a unified form and is as such classified based on its general meaning and grammatical characteristics. Following this, Vietnamese word are divided into either "lexical words" or "form words", with the two categories being comparable to content words and function words respectively.

To avoid the confusion on the classification criteria, we have decided to analyze our corpus and determine POS elements based on lexical- grammatical category, and following the POS classification conducted by the Committee of Social Science (1993). According to the Committee of Social Science, "parts of speech include words with the same general meaning and grammatical characteristics […] The general meaning of Vietnamese words are reflected in their grammatical characteristics. However, their characteristics, in such an isolating language like Vietnamese, are not shown in the phonology but their collocations with other words" (Vietnam Committee of Social Science, 1993, p. 66).

In this classification, lexical words convey the "real meaning" or the "lexical meaning" of objects, and point at the phenomena which establishes the connection between words and objects. In terms of grammar, lexical words can work as "theme" or "rheme" in a sentence. With two lexical words, it is absolutely possible to make a simple sentence.

(1) **Xe** *chạy*
'**Cars** *are moving.*'

(2) **Lúa** *tốt.*
'**The rice** *is growing well*'

On the other hand, form words in Vietnamese do not convey any real meanings, and do not connect to any objects or phenomena. These words themselves cannot function as main parts of a sentence, but have to go with lexical words to make a sentence; hence, they convey grammatical meaning such as time (example (3)) or degree (example (4)).

(3) Xe ***đã*** chạy.
'Cars ***have*** gone.'

(4) Lúa ***rất*** tốt.
'The rice is growing ***very*** well.'

Furthermore, form words can carry additional meanings.

(5) Lúa mùa ***và*** lúa chiêm đều rất tốt.
'The winter rice ***and*** the summer rice grew very well.'

(6) Lúa ***của*** hợp tác xã đó tốt.
'The rice ***of*** that cooperative grew well.'

In order to make the classification more effective and useful in forming sentences, lexical words and form words are divided further into two groups. Lexical words are categorized into nouns, verbs and adjectives; whereas form words are classified into adjuncts and conjunctions. In addition to these categories, we also make the use of pronouns, while modifiers, and interjections are the two categories that belong to both lexical words or form words, and differ from the category of pronouns. To sum up, part of speech in Vietnamese are categorized into eight main groups, of which former six groups are subdivided as follows [2]:

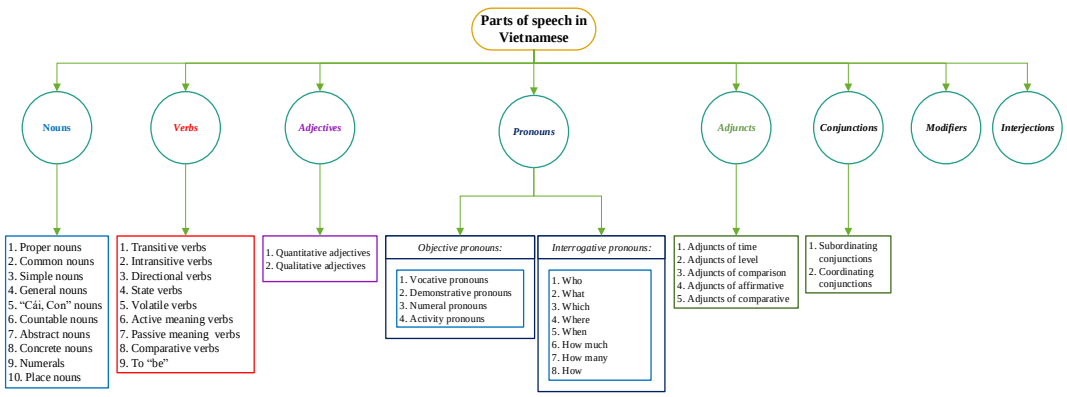


Figure 1: Parts of speech in Vietnamese

3.2.2 POS elements affecting text readability in Vietnamese

Word class is a hierarchical system in which a category consists of smaller categories. Vietnamese words can be divided into the two main categories, cf. lexical words and form words. Each category can be devided further based on the parts of speech. This significance covers a narrower scope of a word, but the meaning remains the general syntactic meaning (Mostafa & Pooneh, 2012, p. 270). Lijun, Martin, Matt and Noemie (2010, re-extracted from Heliman et al. (2007) and Leory et al. (2008)) show that the characteristics of the part of speech in a text prove very useful in determining text readability.

[2] Categorized according to Vietnamese Grammar (1993, pp. 67–95).

To determine the influences of the POS on readability in Vietnamese texts, we used the automatic supporting tool called CLC-Vietnamese-Toolkit, through which we identified 25 common parts of speech in Vietnamese. There were few cases where identification was impossible, and such words were labeled X (unidentified POS - Unknown). We could then investigate the relationship among different parts of speech in text readability, and labelled them with different grade levels (from grade 2 to grade 5).

The corpus was analyzed to determine the frequency of the parts of speech used in each text of each grade. The data showed that some parts of speech were not used at all, and hence the lowest frequency is recorded is zero (0). For example, examining 67 texts in grade 2, we found out that proper noun was not used in 20 of the 67 texts, and 22 times is the largest frequency with which this part of speech was used in texts. Therefore the frequency of proper nouns in grade 2 ranges from from the lowest (0) to the highest (22) as we can see from the extracted data in the Table 2 below:

Table 2: The extracted data of parts of speech in Vietnamese primary textbooks

	A	B	C	D	E	F	G	H	I
1	File	Lớp	Pd	Nu	An	PU	ID	Nc	FW
2	2\Tap 1\0 - Ban cho.txt.ws.pos	2	2	3	0	30	0	6	0
3	2\Tap 1\0 - Ban tay diu dang.txt.ws.pos	2	0	0	0	28	0	10	0
4	2\Tap 1\0 - Be Hoa.txt.ws.pos	2	2	0	0	30	0	9	1
5	2\Tap 1\0 - Can voi.txt.ws.pos	2	2	1	0	9	0	3	0
6	2\Tap 1\0 - Cay xoai cua ong em.txt.ws.pos	2	2	0	0	21	0	0	0
7	2\Tap 1\0 - Di cho.txt.ws.pos	2	6	4	0	22	1	0	0
8	2\Tap 1\0 - Dien thoai.txt.ws.pos	2	2	0	1	32	0	4	2
9	2\Tap 1\0 - Doi ban.txt.ws.pos	2	1	0	0	18	0	1	0
10	2\Tap 1\0 - Doi giay.txt.ws.pos	2	2	0	0	21	0	0	1
11	2\Tap 1\0 - Ga ti te voi ga.txt.ws.pos	2	1	0	0	47	0	0	3
12	2\Tap 1\0 - Ha mieng cho sung.txt.ws.pos	2	1	0	0	19	0	0	0
13	2\Tap 1\0 - Lam viec that la vui.txt.ws.pos	2	1	0	0	29	0	0	1
14	2\Tap 1\0 - Mit lam thu.txt.ws.pos	2	6	1	1	57	0	16	1
15	2\Tap 1\0 - Mua kinh.txt.ws.pos	2	1	0	0	20	0	0	0
16	2\Tap 1\0 - Ngoi truong moi.txt.ws.pos	2	0	0	0	17	0	0	0
17	2\Tap 1\0 - Qua cua bo.txt.ws.pos	2	2	0	0	23	0	0	4
18	2\Tap 1\0 - Them sung cho ngua.txt.ws.pos	2	3	0	0	36	0	2	2
19	2\Tap 1\0 - Tren chiec be.txt.ws.pos	2	0	1	0	20	0	0	0
20	2\Tap 1\1 - Co cung mai sat co ngay nen kim.txt.w	2	1	0	0	27	0	0	0
21	2\Tap 1\10 - Ba chau.txt.ws.pos	2	3	1	0	42	0	0	2
22	2\Tap 1\11 - Su tich cay vu sua.txt.ws.pos	2	2	0	0	42	0	1	2
23	2\Tap 1\12 - Bong hoa Niem Vui.txt.ws.pos	2	2	0	0	29	0	10	1
24	2\Tap 1\13 - Cau chuyen bo dua.txt.ws.pos	2	2	0	0	38	0	0	1
25	2\Tap 1\14 - Hai anh em.txt.ws.pos	2	6	0	0	30	0	0	0
26	2\Tap 1\15 - Con cho nha hang xom.txt.ws.pos	2	2	0	0	39	0	22	1
27	2\Tap 1\16 - Tim ngoc.txt.ws.pos	2	4	0	0	42	0	1	0
28	2\Tap 1\2 - Phan thuong.txt.ws.pos	2	3	0	0	31	0	4	0
29	2\Tap 1\3 - Ban cua Nai Nho.txt.ws.pos	2	3	0	0	29	0	10	2
30	2\Tap 1\4 - Bim toc duoi sam.txt.ws.pos	2	2	0	0	42	0	13	1

Using the CLC-Vietnamese-Toolkit, we examine the parts of speech of the texts in each grade. Based on the statistics, their frequency was calculated, and results are listed in Table 3:

Table 3: The frequency of the POS elements affecting text readability in Vietnamese - prose corpus, primary textbooks

No.	Part of speech	POS	Grade 2	Grade 3	Grade 4	Grade 5	Total
1	Proper Nouns	Nr	0–22	0–20	0–24	0–23	0–24
2	Countable Nouns	Nc	0–16	1–19	0–34	1–15	0–34
3	Concrete Nouns	Nu	0–4	0–10	0–8	0–4	0–10
4	Temporal Nouns	Nt	0–19	0–22	0–19	0–14	0–22
5	Numerals	Nq	1–18	2–25	3–29	4–24	1–29
6	Common Nouns	Nn	11–83	25–89	38–146	28–107	11–83
7	Directional Verbs	Vd	0–6	0–10	0–8	0–5	0–10
8	State Verbs	Ve	0–11	0–8	0–8	0–10	0–11
9	Comparative Verbs	Vc	0–8	0–6	0–6	0–7	0–8
10	Volatile Verbs	Vv	12–74	17–72	17–104	19–90	12–104
11	Directions	D	0–8	0–11	0–9	0–10	0–11
12	Quantity Adjectives	An	0–2	0–5	0–8	0–4	0–8
13	Quality Adjectives	Aa	1–24	4–33	8–43	6–39	1–43
14	Demonstrative Pronouns	Pd	0–8	0–7	0–12	0–11	0–12
15	Personal Pronouns	Pp	0–23	0–38	0–39	0–33	0–39
16	Adverbs	R	1–31	1–33	3–51	2–45	1–51
17	Prepositions	Cm	1–18	1–29	1–29	3–19	1–29
18	Parallel Conjunctions	Cp	0–17	1–17	4–33	3–22	0–33
19	Subordinating Conjunctions	Cs	0–4	0–4	0–3	0–8	0–8
20	Modifiers	M	0–10	0–9	0–12	0–6	0–12
21	Emotion Words	E	0–4	0–4	0–3	0–2	0–4
22	Foreign Words	FW	0–5	0–7	0–6	0–6	0–7
23	Onomatopoeia	ON	0–0	0–2	0–0	0–0	0–2
24	Idioms	ID	0–1	0–1	0–2	0–1	0–2
25	Unidentified POS	X	0	0	0	0	0

According to the corpus analysis outlined above, we can first quantitatively identify 25 elements which affect text readability. The frequency of each element for each grade (from 2 to 5) and the elementary level are identified. For example, the frequency of "proper nouns" in a text of elementary level, from grade 2 to 5, is from 0 to 24, more specifically, in Grade 2, the frequency is from 0 to 22, 0 to 20 for Grade 3, 0 to 24 for Grade 4, and 0 to 23 for Grade 5. The frequency of elements from other categories can also be identified in a similar way. In each grade, 25 parts of speech can be determined in their scopes, out of which we can see differences among language elements per grade as well as per all grades together.

It can be seen from Table 3 that no text at elementary level are uses unidentified POS, and hence investigating this linguistic element in Vietnamese texts at

intermediate and advanced levels is necessary for robust conclusions. The rest 24 elements from the remaining types can be classified into 3 groups: the frequency at low levels (0-34), the frequency of average (35-68) and the group with a high level of frequency (68-104). This is shown in Table 4 below:

Table 4: POS Elements affecting text readability in Vietnamese – Elementary level

POS Elements affecting text readability in Vietnamese - elementary level		
Elements with low frequency	Elements with average frequency	Elements with high frequency
Proper Nouns	Qualitative Adjectives	Common Nouns
Countable Nouns	Personal Pronouns	Volatile Verbs
Concrete Nouns	Adverbs	
Temporal Nouns		
Numerals		
Directional Verbs		
State Verbs		
Comparative Verbs		
Directions		
Quantitative Adjectives		
Demonstrative Pronouns		
Prepositions		
Subordinating Conjunctions		
Parallel Conjunctions		
Modifiers		
Emotion Words		
Foreign Words		
Onomatopoeia		
Idioms		

Besides investigating the POS frequency, we also examine the correlation of these elements to determine their influences on Vietnamese text readability. We used Pearson Correlation to compute these numbers.[3] In this way, we examined linear relations between the POS elements (independent variables) and Vietnamese text readability (dependent variable) by Pearson correlation coefficient (depicted by r). The value of the correlation coefficient ranges from -1 to 1, with r = 0 (or close to 0) suggesting that there is no or very weak relation between a POS element (x) and Vietnamese text readability (y). In cases when correlation coefficient ranges below 0 ($r < 0$), the two correlate inversely, namely that x increases with the decrease of y and

[3] http://phantichspss.com/he-so-tuong-quan-pearson-cach-thao-tac-phan-tich-tuong-quan-trong-spss.html

the other way around. And finally, in cases when correlation coefficient ranges above 0 ($r < 0$), the two correlates show direct relation; when x increases, y will increase. The correlation analysis results are presented in Table 5.

Table 5: The Pearson correlation between the POS elements and Text readability

Part of speech	*r*	Part of speech	*r*
Demonstrative Pronouns	0.160	**Emotion Words**	**-0.111**
Concrete Nouns	0.167	Countable Nouns	0.206
Quantity Adjectives	0.098	**Common Nouns**	**0.511**
Idioms	0.071	Quality Adjectives	0.443
Proper Nouns	0.232	Numerals	0.355
Foreign Words	0.142	Personal Pronouns	0.017
Directional Verbs	0.052	Adverbs	0.231
Volatile Verbs	0.351	**Onomatopoeia**	**-0.026**
Comparative Verbs	0.255	Modifiers	0.019
Prepositions	**0.509**	Coordinating Conjunctions	0.402
Directionals	0.102	State Verbs	0.207
Temporal Nouns	0.229	Subordinating Conjunctions	0.115

The correlation analysis results show that most of them are positively related; and there are only two negative correlation coefficients with text readability: emotion words (-0.111) and onomatopoeia (-0.026); but the influence of two elements on text readability is relatively low (nearly no affection). Among 22 POS elements with positive correlation coefficients, frequencies of common nouns and prepositions have strong connection with the text readability (0.511 and 0.509). This in other words means that in case of common nouns, about 30% of the change of text readability links to the change in frequency of other nouns in the texts. Similarly, the correlation coefficient of prepositions means that, with all the elements being analyzed, about 26% of the change of text readability is related to the change of the frequency of prepositions.

From the above results we can suggest that the two POS elements, namely prepositions and common nouns are the most influential linguistic elements in Vietnamese text readability, such as polysemantic common nouns or prepositions in ambiguity. AS such they are expected to gain attention in further studies.

4 Comments and conclusion

The survey about the extent to which 25 POS elements affect text readability in prose texts in Vietnamese textbooks for primary pupils at elementary level (easy) can help teachers, editors, and learners to determine the level of difficulty qualitatively. The findings, in this level, show that common nouns and volatile verbs are the elements

with the highest frequency, three of the parts of speech with the medium frequency are qualitative adjectives, personal pronouns, and adverbs. Except for the unidentified POS, the rest of the parts of speech - 19 categories- are used with low frequency. In addition, the correlation coefficient also shows that conjunctions and common nouns are the potential language elements affecting Vietnamese text readability, and their meaning and grammatical structure should be investigated further.

The most important thing in evaluating POS elements affecting text readability is that the corpus must be classified in different levels. However, at present, there is no tool or formula reliable or effective enough to measure the text readability for Vietnamese texts. Therefore, choosing a corpus collected from the textbooks which were already classified into different grade levels for elementary school- aged children is ideal for this study. Besides, there are still some issues in the corpus itself. Although the texts hierarchically divided in increasing levels from Grade 2 to Grade 5, there is no clear distinction. For example, the frequency of temporal nouns in grade 2 and 4 is equal (0 -14), while grade 3 has the highest frequency (0-22) and 5th grade, despite being the highest grade, has the lowest frequency (0-14). Therefore, further studies with a larger corpus for this level as well as in intermediate and advanced levels are necessary.

Text readability in English has been studied since the early 19th century, but investigating text readability in Vietnamese is still the beginning. Therefore, in the future, we will build a larger corpus from multiple materials as well as divide the corpus using both quantitative and qualitative methods to calculate at three levels: basic; intermediate; and advanced. We will also investigate Vietnamese readability more deeply with other linguistic elements. Since then, the analysis of the corpus is more reliable and convincing. It can help the computational linguistics to build applicable formula or tools for measuring text readability for Vietnamese, a low- resourced language, to meet the demand for users and Vietnamese community in this era of technology.

References

Bùi, M. H. (2008). *Ngôn ngữ học Đối chiếu*. Hồ Chí Minh City, HCMC: Education Publishing House.

Cao, X. H., & Hoàng, D. (2005). *Từ Điển Thuật ngữ Ngôn ngữ học Đối chiếu Anh - Việt; Việt – Anh*. Hồ Chí Minh City, HCMC: Social Sciences Publishing House.

Cieri, C., Maxwell, M., Strassel, S., & Tracey, J. (n.d.). *Selection Criteria for Low Resource Language*. University of Maryland College Park, MD 20742, USA. Retrieved from https://pdfs.semanticscholar.org/315a/3a4a6db25e705f50159807917ec6f439f83b.pdf

Council of Euroupe (2010). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge Press. Retrievd from http://www.coe.int/en/t/dg4/linguistic/source/framework_en.pdf

Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English, 26*, 23.

Dubay, H. W. (2004). *The Principles of Readability.* Impact Information, Costa Mesa, California.

Đinh, Đ. (2006). *Xử lý Ngôn ngữ tự nhiên*. Hồ Chí Minh City, HCMC: HCMC National University Publishing House.

Flesch, R. (1943). *Marks of a readable style, Columbia University contributions to education*, no. 897. New York: Bureau of Publications, Teachers College, Columbia University.

Flesch, R. F. (1948). A New Readability Yardstick. *Journal of Applied Psychology, 32*(3).

Flesch, R. F. (1949). *The Art of Readable Writing.* New York: Harper.

Flesch, R. F. (1974). *The Art of Readable Writing,* 25th anniversary edition, revised and enlarged. NY: Harper & Row.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh- Metrix: Providing Multilevel Analyses of Text Characteristics, *Educational Researcher*, *40*(5), 223-234.

Graesser, A.C., McNamara, D. S., & Louwerse, M. M. (2004). Coh- Metrix: Analysis of Text on Cohesion and Language, *Behavior Research Methods, Instruments, & Computers*, *36*(2). 193-202.

Gray, W. S., & Leary, B. E. (1935). *What Makes a Book Readable*. Chicago, Illinois: The University of Chicago Press.

Jamie, D. (2014). *Investigating the relationship between empirical task difficulty, textual features and CEFR levels*. EALTA 2014, 29 May – 1 June. University of Warwick

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. *CNTECHTRA Research Branch Report 8-75*.

Klare, G. R. (1973). *The Measurement of Readability*. Ames, Iowa: Iowa State University Press.

Lijun, F., Martin, J., Matt, H., & Noémie, E. (2010). *A comparision of Features for Automatic Readabiity Assesment,* Beijing August 2010, Poster Volume, 276-284.

Lorge, I. (1939). Predicting Reading Difficulty of Selections for Children. *The Elementary English Review*, *16*(6), 229-233.

Mai, N. C., Nguyễn Thị, N. H., Đỗ, V. N., & Bùi, M. T. (2007). *Nhập môn Ngôn ngữ học*. Hồ Chí Minh City, HCMC: Education Publishing House.

McLaughlin, H. (1969). SMOG Grading - a New Readability Formula. *Journal of Reading, 12*(8), 639-646. DOI: http://dx.doi.org/10.2307/40011226

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*, CUP.

Mostafa, Z., & Pooneh, H. (2012). Readability of Texts: State of the Art. *Theory and Practice in Language Studies*, *2*(1), 43-53. http://doi.org/10.4304/tpls.2.1.43-53.

Nguyễn, T. G., Đoàn, T. T., & Nguyễn, M. T. (2008). *Dẫn luận Ngôn ngữ học*. Hồ Chí Minh City, HCMC: Education Publishing House.

Viet Nam Committee of Social Sciences. (1993). *Ngữ pháp tiếng Việt*. Hanoi: Social Science Publishing House.

Vietnam Committee of Social Science (1993). *Vietnamese Grammar*. Hanoi: Social Science Publishing House.

Vu Thi, P. A. (n.d.). *Text Readability and testing languages*. Retrieved from http://ncgdvn.blogspot.com/2011/10/o-kho-cua-van-ban-va-viec-kiem-tra-ngon.html

Vu Thi, P. A. (2006). Khung trình độ chung Châu Âu và việc nâng cao hiệu quả đào tạo tiếng Anh tại ĐHQG – HCM. *Journal of Science and Technology Development*, *9*(10), 31-47.

Appendix

The parts of speech in Vietnamese calculated by CLC - Vietnamese – Tool kit, Computing Linguistics- CLC- University of Science Ho Chi Minh City

Label	Từ loại tiếng Việt (Vietmamese POS)	Ví dụ (Example)	Từ loại tiếng Anh (English equivalents)
Aa	tính từ hàm chất	lộ thiên, đầy, mắc	qualitative adjective
An	tính từ hàm lượng	đầu tiên	quantitative adjective
Cm	giới từ	giữa, của, trong, tại	major/minor conjunction
Cp	kết từ đẳng lập	cùng, với, và	parallel conjunction
Cs	kết từ chính phụ	nếu, thì, vừa, là	subordinating conjunction
D	phó động từ chỉ hướng	ra, vô, lên, xuống	direction
E	cảm từ	thưa, làm gì	emotion word
FW	từ nước ngoài	Miss, pH, super	foreign words
ID	thành ngữ	công ăn việc làm	idiom
M	trợ từ	đến, riêng, được, có, đó	modifier
Nc	danh từ đơn thể	bộ, ngôi, bản, con, bài	countable noun
Nn	danh từ	nước, người, chuyện, ông	common noun
Nq	danh từ số lượng	một vài, phần lớn, mấy	numeral
Nr	danh từ riêng	Tuấn, Hồng, Thành, Hà Nội	proper noun
Nt	danh từ chỉ thời gian	sáng, tối, năm, khi	temporal noun
Nu	danh từ chỉ đơn vị	TP., tỉnh, khu phố	concrete noun
ON	từ tượng thanh	tách, bùm bụp, hì hì	onomatopoeia
Pd	đại từ không gian, thời gian	nào, này, đó, bao giờ	demonstrative pronoun
Pp	đại từ xưng hô	tui, con, anh, chị, ông	personal pronoun
PU	dấu câu	Dấu phẩy, dấu chấm	punctuation
R	trạng từ	được, đều, chưa, nào	adverb
Vc	động từ so sánh	Là	comparative verb
Vd	động tử chỉ hướng	đến, ra, xuống	directional verb
Ve	động từ tồn tại	có, hết	state verb
Vv	động từ ý chí	viết, muốn, được, thay, ăn	volatile verb
X	không xác định	v.v	unidentified POS