Spinning a yarn across languages: Adapting MAIN for India

Manish Madappa

The English & Foreign Languages University, Hyderabad

Yozna Gurung

The English & Foreign Languages University, Hyderabad

Madhavi Gayathri Raman

The English & Foreign Languages University, Hyderabad

The Multilingual Assessment Instrument for Narratives (LITMUS-MAIN) was developed to assess the narrative abilities of bi- and multilingual children in the various languages that they speak. This paper presents the details of the adaptation of MAIN to three Indian languages, Kannada, Hindi and Malayalam. We describe some typological features of these languages and discuss the challenges faced during the process of adaptation. Finally, we give an overview of results for narrative comprehension and production from Kannada-English and Hindi-English bilinguals aged 7 to 9.

---

1   Introduction

The linguistic landscape of India is a complex one given the pluri-lingual and pluri-ethnic nature of the country. The Constitution of India recognizes Hindi and English as official languages of the country. The 2011 Census of India lists a total of 121 languages of which 22 are Scheduled languages, meaning that they are included in the Eighth Schedule of the Indian Constitution and given official status, and 99 are Non-Scheduled languages. Each State and Union Territory has its own official language. English plays an important role in India due to its colonial past. Used initially for administrative purposes, over time, English has become the *lingua franca* that links people of different linguistic communities for social, educational, political, and economic purposes.

The 2011 Census of India reports that 250 million Indians speak at least two languages and more than 85 million speak three or more languages. Thus, it is evident that in India bilingualism or multilingualism is the norm. One of the greatest challenges facing researchers and educators in such a context is assessing the linguistic proficiency of bilingual children in all of the languages that they speak. This has been particularly difficult because there have been very few assessment tools that have been developed specifically for bilinguals. Given the wide number of languages that exist and the differences between languages, it is difficult to arrive at comparable measures across the various components of a language using a single assessment tool. The use of narratives and the universality of narrative abilities across languages allows for the assessment of multiple features of language in all of the languages of bi- or multi-linguals.

The *Language Impairment Testing in Multilingual Settings – Multilingual Assessment Instrument for Narratives* (LITMUS-MAIN, hereafter MAIN; Gagarina et al., 2012; 2015; 2019) is a tool that has been developed to assess the narrative skills of young bilingual and multilingual children between the ages of 3 and 10 years, but has also been used to assess narrative skills of older bilinguals and adults as a control group (e.g. Gagarina, Bohnacker & Lindgren, 2019). MAIN consists of four picture stories each consisting of the six sequences of pictures. These stories have been carefully matched such that they are comparable in terms of the components of Macrostructure, Microstructure and Internal State Terms. The instrument assesses both comprehension and production of narratives in all the languages that a child speaks. Since it has been developed specifically for bilinguals it is less biased towards bilingual populations with diverse language and cultural backgrounds than other assessment tools norm referenced for monolingual populations. This provides the rationale for adapting MAIN to Indian languages. This paper describes the adaption of MAIN to Hindi, Kannada and Malayalam.

2 Brief descriptions of the three languages

This section provides brief descriptions of the three Indian languages, Hindi, Kannada and Malayalam, to which MAIN has been adapted by the authors of this paper.

2.1 *Hindi*

Hindi, a language spoken in the northern part of India, belongs to the Indo-Aryan group of languages within the Indo-Iranian branch of the Indo-European language family. There are approximately 420 million speakers of Hindi in the world. In India, Hindi is spoken primarily in the states of Bihar, Chattisgarh, Delhi, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttarakhand, and Uttar Pradesh. Along with English, Hindi is also the official language of India. Hindi is written using the Devnagari script and each alphabet has a one to one correlation with the phoneme inventory (high level of sound-symbol correspondence). Though Hindi is essentially a verb-final, SOV, language, as shown in (1), it permits other word orders as well (Kidwai, 2008).

(1) m̃ene   patr   likha:
     I-ERG   letter wrote-PAST
     'I wrote a letter.'

Nouns in Hindi are inflected for gender, number, and case. The gender of the words for inanimate objects cannot be predicted from their form or meaning. Pronouns are inflected for number and case. In contrast to English prepositions, Hindi uses postpositions which follow the words they govern. They are written as separate words with nouns but are tagged to pronouns (Koul, 2008). Verbs are inflected for person, number, gender, tense, mood, and aspect. The negation is placed before the verb. The verb agrees with the subject for number and gender, as in (2) and (3).

(2) larka   ghar   ja:ta: hɛ.
     the boy   home   go-VB-MASC   is/be
     'The boy goes home.'
(3) larki: ghar   ja:ti: hɛ.
     the girl   home   go-VB-FEM   is/be
     'The girl goes home.'

Hindi uses a number of light verb constructions as well and these "light verbs have been assumed to be responsible for hosting tense and aspect features, licensing arguments, and functioning as auxiliarie/s" (Suleman, 2015, p. 3). An example is given in (4).

(4) vah   kha:na: kha: chuka: hɛ.
     he   food   eat   PERF-MSC   is/be.
     'He has eaten his food (completion).'

2.2   *Kannada*

Kannada is a Dravidian language. Dravidian languages are spoken mainly in southern India and south Asia and are divided into four main groups: South, South-Central, Central, and North. Of the 26 languages that belong to these groups, Tamil, Telugu, Kannada, and Malayalam belong to the South Dravidian group (Amritavalli, 2017). They have a history of literary traditions. According to the 2011 census, there are more than 215 million speakers of Dravidian languages in India (Census, 2011). Kannada is spoken mainly in the state of Karnataka and by linguistic minorities in the neighbouring states of Maharashtra, Andhra Pradesh, Tamil Nadu and Kerala. It has approximately 44 million native speakers (Census, 2011) and is considered the third oldest Indian language after Sanskrit and Tamil. Roughly 12.9 million speakers in Karnataka use Kannada as a second or third language. It is the official language as well as the language of administration of Karnataka. Kannada, written using the Kannada script is an alphasyllabic language, i.e., it represents language at roughly the level of the syllable, but its symbols or

'akshara' can be segmented visually to reveal the consonant and vowel segments within (Nag, Treiman & Snowling, 2010).

Though the preferred word order is verb-final, SOV, other word orders are also permitted so that the word order of Kannada can be said to be relatively free. Being an agglutinative language, Kannada forms words by the addition of suffixes onto the root word. Nouns and pronouns in Kannada are marked for gender, case and number. There are two natural genders, masculine and feminine; all inanimate objects belong to the neuter gender. Like Hindi, Kannada uses postpositions that are added to nouns phrases after the case marker, as in (5).

(5)   mane   hinde
        house   behind
        'behind the house'

Kannada verbs are either finite or non-finite with finite verbs placed at the end of a sentence, as shown in (6).

(6)   avrig   erD   makL   iddaare
        to him   two   children   exist
        'He has two children.'

2.3   *Malayalam*

Malayalam, a member of the South Dravidian subgroup of the Dravidian language family, is the official language of the southern-most state of India, Kerala, as well as the union territories of Lakshadweep and Puducherry (Pondicherry). Malayalam is also spoken by a significant number of linguistic minorities in the neighbouring states of Karnataka and Tamil Nadu. It is also widely spoken in Gulf countries due to the presence of a large number of Malayali expatriates. It is one of the 22 official languages of India and is spoken by nearly 37 million people as a first language and by about 700,000 as a second language. Malayalam is written using a Brahmic script which is an alphasyllabary (or *abugida*, a writing system where the consonant letters represent syllables with a default vowel and other vowels are denoted by diacritics) like the Kannada script. The most common word order is verb-final, SOV but other word orders are also possible. Malayalam is an agglutinative language and suffixes are added to noun and verb stems to mark grammatical categories. Since there is no fixed limit on the number of suffixes that can be added, it often leads to the formation of long words, as shown in (7).

(7)   suṇḍaɾ-an-maaɾ-ey-um   suṇḍaɾ-i-maaɾ-ey-um

        beauty-M-PL-ACC-CORD   beauty-F-PL-ACC-CORD

        'Handsome men and beautiful women.'
        (Gayathri, 2019)

The verb in Malayalam is not marked for either gender or number, as in (8). Malayalam verbs do not exhibit concord with noun phrases that are their arguments. Neither subject, direct object, nor indirect object are coded on the verb (Ashrer & Kumari, 1997, p. 348). On nouns, the suffix –an marks masculine gender and the suffix –i, feminine gender, as in (9).

(8) aval nalla kuṭṭi aaṇə
    she good child be-PRES
    'She is a good girl.'   (Asher & Kumari, 1997)

(9) kaḷḷan - kaḷḷi 'Thief'   (Nair, 2012)
     miṭukkan - miṭukki 'Smart person'

It is possible to delete the verb in some types of copula constructions where a time reference is not required. Unlike Hindi and Kannada, gender in Malayalam is indicated in the form of a noun that expresses natural gender or is marked on 3rd person and demonstrative pronouns. Like Kannada and Hindi, Malayalam is a postpositional language, as shown in Table 1.

Table 1. Comparison between Hindi, Kannada and Malayalam

Language	**Example**			
Hindi	pahar	se	nadi:	tak
Kannada	parvata	dinda	nadiya	varege
Malayalam	mala	mutal	puza	vare
	mountain	from	river	up to
'from the mountain up to the river' (Asher & Kumari, 1997)				

3   Challenges during the adaptation of MAIN to Kannada, Hindi and Malayalam

The process of adapting MAIN into the three Indian languages described above was a complex one. Since these languages belong to language families that are typologically different from English, the task of matching the various elements of macrostructure (story grammar, structural complexity, and internal state terms), and microstructure (total number of words used) was challenging. For instance, the realization of complete sentences in English contained either fewer or more arguments (in terms of the number of words and clausal constituents) in these languages. Consider for example, the sentence in Kannada in (10).

(10)   taayiyu nodi-tu   marigalu   hasivu-agigdd-annu
       mother see-PAST children   hunger-ACC-PAST
       'The mother saw that the baby birds were hungry.'

Given the syntax of Kannada, it was not always possible to translate the story scripts from English into Kannada by the use of the same syntactic structure and maintain the same order of events within one sentence. In a number of cases, the English sentence has to be split into two.

For example, the *Baby Birds* story begins with 'One day there was a mother bird who saw that…' This, when translated into Indian languages has to be split into two sentences to sound natural. Alternatively, a change in the order of events becomes inevitable. This is the case with the other stories as well. The agglutinative nature of Kannada and Malayalam which are morphologically rich languages posed challenges since different word forms are formed by inserting different morphemes to the root word serially. For example, as in (11) and (12):

(11)   apāya-dallida-dannu (Kannada)
        danger-PAST-in
        'were in danger'
(12)   viśakkunn-uṇṭāy-irunnu (Malayalam)
        hungry-PROG-be-PAST
        'was hungry'

In all three languages, the choice of vocabulary proved to be a problematic aspect. Several words or phrases which do exist in these languages are seldom used in speech. For example, in Kannada, equivalents semantically close to English like *dhavisu* 'leaped' or *negeyithu* 'jumped' are hardly used in spoken discourse. Instead, words like *dikkihodi* or *guddiko* 'bumped' are used in the same context. There is a slight difference in meaning; *gikkihodi* suggests that the person or thing intentionally or deliberately bumped into something. Since most of the participants (see Section 4) used these words in the narratives, they were included in the scoring sheets.

As a result of the ubiquitous presence of English in the environment,[1] many words in Indian languages have been replaced by their English equivalents in text and in speech. For example, the Kannada words, *chendu* and *pantu* have been replaced by *ballu*, i.e., the English word *ball* with the addition of the *-u* suffix. The pictures in the *Dog* story depicted an unfamiliar food item 'sausage' which is culturally alien and does not have an equivalent in Indian languages. These were replaced with more familiar images of *mansa* 'meat'. Later, after having tested a number of children (see Section 4), these images were modified further and a special version of the *Dog* story was created for India in which the food in the *Dog* story is popular vegetarian snack. The scoring protocols were also modified to include these changes.

In order to ensure that the vocabulary and the structures used in the story scripts were age appropriate, school textbooks were used as a resource. The opening sentence of each of the stories 'Once upon a time…' can be realized in at least three different ways in the Indian story telling tradition. For example, in Hindi, it is possible to begin the story with *ek baar ki baat hain* 'once upon a time' (lit. 'once time talk is'), *bahut pehele ki baat hain* 'a long time ago' (lit. 'very long before talk is') or *ek din ek…* 'once there was' (lit. one day, one…). All of these are equally valid and acceptable ways of beginning a story and all of them were accepted in the narratives produced by the children.

---

[1] English is a second language in India and is used throughout the country for a variety of purposes. It is the medium of instruction in many schools. In schools where instruction is imparted in the mother tongue, English is taught as a subject either from primary or middle school. It is the main medium of instruction in higher education.

The adaptation of the instrument into Kannada was led by the first author, a native-speaker of the language, while the second author carried out the Hindi adaptation under the guidance of the third author. The Malayalam adaptation was carried out by the third author, a native-speaker of that language. Each of the three language versions were then sent to two linguists each who are native speakers of those languages. Back translation was done and checked. Finally, three adult native speakers with college-level education also proof-read the instrument.

The Kannada MAIN version was then piloted with 10 Kannada-English middle-income SES bilinguals studying in an English-medium school in Mysore, Karnataka. They were aged 7;5 to 9 years and were fluent speakers of Kannada. The Hindi version of MAIN was piloted with seven children aged 7;8 to 9 years who were enrolled in Kendriya Vidyalaya (one of many nation-wide schools set up for the children of officers of the Central Government including Defence and Para-military personnel), an English-medium school. All students were fluent speakers of their home language, Hindi and their families belonged to the middle-income socio-economic group. The Malayalam version has not yet been piloted.

The initial adaptation of the scripts and protocols used language that was formal and literary and attempted to be a close translation of the English version. After the piloting, we found that the stories sounded more natural when we used spoken language equivalents for words like 'ball' and 'balloon' which was more colloquial and closer to the oral discourse of storytelling. The scripts and protocols were modified so that they reflected everyday child speech.

4 The use of the Hindi and Kannada MAIN

The Kannada and Hindi MAIN-versions and the original English instrument were administered to 31 Kannada-English and 31 Hindi-English bilinguals in Grade 3. The Kannada-English group was between 7;5 and 9 years old (mean age = 8;1) and the Hindi-English group was between 7;8 and 9 years old (mean age = 8;3). The children were recruited from the same schools where the pilot studies were conducted.

The idea of telling stories based on pictures appealed to the children and they performed the task enthusiastically. However, the *Dog* story proved particularly difficult for the children in both languages in two respects: (1) they had difficulties expressing the concept 'bumped into the tree'; and (2) the replacement of sausages with meat in the pictures did not make it easier for the children as we had expected. For this reason, the version of the MAIN pictures to use with these groups now contains a popular vegetarian snack instead of sausages or meat.

The results from the *Hindi-English bilinguals* revealed that the scores obtained on story structure in Hindi and English were comparable. The Hindi-English bilinguals included on average 44.78% of the 17 MAIN story structure (macrostructural) components (M = 7.61, SD = 2.7) in Hindi and 43.83% in English (M = 7.45, SD = 3.73). A t-test showed no significant difference (p = .84) lending support to the findings in the literature that narrative Macrostructure is invariant across languages (e.g., Berman, 2001; Fiestas & Pena, 2004; Iluz-Cohen & Walters,

2012). GAO-sequences (complete episodes) were produced in 11.82% and 10.75% out of all episodes for Hindi and English, respectively; this difference is not statistically significant (p = .82). Attempts (A) and Outcomes (O) were most frequently included in both languages, though there was a considerable difference across the two languages for how often these components were produced in an episode: 81.72% (A) and 67.74% (O) out of all episodes in Hindi and 66.66% (A) and 48.38% (O) out of all episodes in English. Goals (G) were produced in approximately 20% of the episodes in both languages. The percentage of correct responses to the comprehension questions in Hindi was 82.25% while in English it was 70.32%. Comprehension was thus much better than production in both languages with an advantage for the first language, Hindi in comprehension. The children's comprehension of Goals was considerably higher in than in their production in both Hindi (87.09%) and English (66.66%).

The results from the *Kannada-English bilinguals* show a difference in the production of narrative macrostructure in Kannada and English. Kannada-English bilinguals included 58.06% (M = 9.87, SD = 2.47) of the 17 MAIN story structure components in Hindi and 50.8% (M = 8.64, SD = 1.85) in English. A t-test showed that this difference was significant (p = .03), contradicting the largely accepted findings in the literature that macrostructure is invariant across languages. GAO-sequences were produced in 26.34 % out of all episodes in Kannada and in only 14.24% out of all episodes in English and this difference was also statistically significant (p < .01). Attempts and Outcomes were produced most frequently in both Kannada and English; 70.43% (A) and 86.02% (O) in Kannada and 69.25% (A) and 79.03% (O) in English. Goals were produced less frequently, in 45.43% out of all episodes in Kannada and in 31.72% out of all episodes in English. The percentage of correct responses to the comprehension questions was 77.74% in Kannada and 74.67% in English. Once again, we find that comprehension is better than production in both languages. Goals were comprehended almost at ceiling in both Kannada (96.24 %) and English (95.16%).

Our results reveal certain similarities and differences between the two groups of bilinguals. The results indicate that while macrostructure production is similar across languages in the Hindi-English group, this was not the case for the Kannada-English group. For both groups and across languages, Attempts and Outcomes are produced at much higher rates than Goals. Overall, comprehension is far ahead of production and comprehension of Goals is better than its production in the narratives.

5   Future plans for MAIN in India

India presents a unique opportunity to expand the scope of MAIN as a tool for the assessment of bilingual and multilingual populations because of the sheer number of languages spoken in the country. The adaptation of MAIN to further Indian languages will address the lack of a standardised assessment tool developed specifically for and norm-referenced with bilinguals. Currently, a team of researchers led by the third author is working towards extending the adaptation of MAIN to the following languages:

- Punjabi: Lillotama Vallecha
- Telugu: Noel Anurag Prashant
- Tamil: Madhavi Gayathri Raman
- Bangla: Tariq Anwar, Rhiddhi Saha, Somak Mandal
- Odia: Jayanta Kumar Das, Subashish Nanda
- Bagri: Radhika
- Konkani: Gautham R Anand, Nagarathna Raveesha
- Nepali: Yozna Gurung

6 Conclusion

The ability to narrate stories is closely linked to later literacy and academic outcomes. MAIN could help teachers develop instructional programmes suited to the needs of their students and would serve as the first step in benchmarking typical bilingual and multilingual language development in languages where such information is not yet available. The process of adaptation and scoring MAIN narratives in Hindi, Kannada and Malayalam brought to the fore certain features of storytelling in the Indian tradition (e.g., concept of time, cyclic progression of events) which are different from the Western narrative tradition suggesting that there might be a need to develop a narrative assessment instrument, or a way of scoring narratives, that is more adapted to the Indian tradition in the future.

7 References

Amritavalli, R. (2017). Morphosyntax of Dravidian Languages. *Oxford Research Encyclopedia, Linguistics*. DOI: 10.1093/acrefore/9780199384655.013.272

Asher, R. E. & Kumari, T. C. (1997). *Malayalam*. New York: Routledge.

Berman, R. A. (2001). Narrative development in multilingual contexts: A cross-linguistic perspective. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative development in a multilingual context* (pp. 419–428). Amsterdam: John Benjamins.

Census (2011). *Languages by state*. Retrieved on 5th May 2019 from censusindia.gov.in.

Fiestas, C. E., & Pena, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, *35*, 155–168.

Gagarina, N., Bohnacker, U., & Lindgren, J. (2019). Macrostructural organization of adults' oral narrative texts. *ZAS Papers in Linguistics*, *62*, 190–208.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*, 1–115.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of Narrative Abilities in Bilingual Children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.),

Assessing multilingual children disentangling bilingualism from language impairment (pp. 243–269). Bristol: Multilingual Matters.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, *63*, 1–36.

Gayathri, G. (2019). *Malayalam Morphosyntax: Inflectional Features and their Acquisition*. Unpublished PhD dissertation. IIT: Bombay.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, *15*, 58–74.

Kidwai, A. (2008). *XP-adjunction in universal grammar: scrambling in Hindi-Urdu*. Oxford: Oxford University Press.

Koul, O. N. (2008). *Modern Hindi Grammar.* Virginia: Dunwoody Press.

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.

Nair, R. S. (2012). *A Grammar of Malayalam*. Retrieved on January 5th from www.languageinindia.com

Suleman, F. (2015). *Light verbs in Hindi*. Master of Social Science dissertation, University of KwaZulu-Natal, Durban.