1. Abstract

This work studies the effects of applying English AMR notations to Hindi sentences selected from the Hindi version of The Little Prince (नन्हा राजकुमार). Interesting results are discovered about the sufficiency / ambiguity of PropBank frames, information loss from disregarding light verbs, effects of Hindi particles, and the consequences of disregarding morphosyntactic information on the predicate and its arguments.

2. Insufficient PropBank frames or more than one disambiguable frame for a lemma

The following examples with the predicate 'आना', to come, include (2) with a sense of 'having knowledge of a skill' that is not in PropBank (Kingsbury & Palmer, 2002) for the lemma 'come'. (1) has the come-01 sense of the lemma.

(1) पर दुर्भाग्यवश यहां कोई **आता** ही नहीं
     But unfortunately here nobody **come.HAB** EMP NEG

(2) मुझे भेड़ बनाना नहीं **आता** था
     I.OBL sheep draw NEG **know.HAB** COP.PAST

Relatedly, the verb predicate 'बनाना' (to make) changes its sense with the object it takes. If its object is a picture, it takes on the sense of 'to draw' and adopts draw-01, as in (4). In another sense, the turn-02 (transformation,becoming) frame is invoked as in (3) which takes an animate object 'you'

(3) अच्छा बेवकूफ **बनाया** मैंने तुझे
     complete idiot **turn.PERF.MASC** I.ERG you.OBL

(4) आखिर मैंने सांप के पेट के अंदर की भी तस्वीर **बनाई**
     finally I.ERG snake POSS stomach POSS inside TOPIC EMP picture **draw.PERF.FEM**

3. Information loss from Compound and Conjunct verbs

Definitions of compound and conjunct verbs follow from (Koul, 2008): compound verbs consist of a main verb predicate followed by a light verb, and conjunct verbs consist of a noun or adjective predicate followed by a light verb. English AMR ignores light verbs (Banarescu et al., 2013), which causes information loss when used on Hindi sentences. A contrastive pair with conjunct verbs is below. The main predicate in both examples is the noun 'याद' or 'memory', which combines with the light verbs **कर** and **आई.**

(5) बाद में मैंने **याद_कर** के उस की जो सबसे अच्छी तस्वीर बनाई वह यह रही
     later ADP I.ERG **recall** CONJ he.OBL POSS REL most good picture draw.PERF it this COP

(6) उसे उन सूर्यास्तों की **याद_आई** जिन्हें वह कुर्सी खिसका-खिसका कर देखता रहता था
     he.DAT those sunsets TOPIC **recall** which he.NOM chair move.REDUPLICATED CONJ see.HAB keep.HAB COP.PAST

(5) is an example of the agentive sense of the verb 'to recall', with the subject specially marked by the agentive postposition ने (Arora et al., 2021). The light verb in (6) delineates the inchoative sense (Koul, 2008); the predicate takes a 'dative subject' i.e. the subject has the act of recalling 'happen to them'. The PropBank frames for both recall (recall-02) and remember (remember-01) take an ARG0 'rememberer' argument, which does not distinguish between agent and experiencer. Loss of the light verbs **कर** and **आई,** along with loss of case information on the subject, has led to loss of the agentive / inchoative distinction using English AMR notation.

Light verbs in compound verb structures contribute an additional meaning that becomes lost with English AMR. As an example, the light verb **पड़** denotes 'action for or towards self' (Koul, 2008). In (8), the extra meaning contributed is that a beetle fell towards me. This extra meaning in (8), which becomes apparent given the context provided by the earlier sentence in the corpus (7), is lost. (7) does not have a gloss and is translated from the English Little Prince except for the missing attribution in parentheses.

(7) व्यवसायी ने सिर उठाया, "मैं चौवन साल से इस ग्रह पर हूं और इस बीच मेरे काम में केवल तीन बार बाधा पड़ी है"

(The businessman lifted his head): "During the fifty - four years that I have inhabited this planet, I have been disturbed only three times"

(8)	पहली	बार,	बाईस	साल	हुए,	एक	भौंरा	न	जाने	कहां	से	गिर	**पड़ा**		था
	first	time,	twenty-two	years	be.PERF,	one	beetle	NEG	know	where	ABL	fall	**<action towards self>**		COP.PAST

A table of the light verbs and their extra meaning is in the Appendix, from (Koul, 2008)

4. Inconsistent treatment of Particles

Hindi particles contribute different meanings in different situations, and consequently it is difficult to treat them consistently with English AMR.

The particle 'वाला' (VALA) is perhaps the most complex. In (9), it combines with the noun बत्ती 'lamp' to denote a person in charge of the lamps. This follows from the dictionary definition: a suffix denoting an agent doer owner possessor keeper or inhabitant (Caturvedi, 1970)

(9)	राजकुमार	को	अपने	कर्तव्य	के_प्रति	इतना	ईमानदार	बत्ती	**वाला**	बड़ा	भला	लगा
	prince	DAT	PRON.REFL	responsibility	towards	such	faithful	lamp	**VALA**	very	good	feel.PERF

It is necessary to keep the suffix in the AMR graph to represent the concept of the 'keeper of the lamps', without which the concept just becomes the noun 'lamp'. It's also unclear what AMR relation to introduce between the noun and the particle; as a first pass, the combination of the noun and particle ('बत्तीवाला') is considered as the single concept of type 'person'. Another possibility is to define the VALA particle as a 'person' concept with a **:mod** relation to the noun; however VALA is not always a person type as seen below, and this treatment leads to inconsistencies.

A possible heuristic to use here is to determine the co-occurrence of the particle with nouns to pinpoint its use as an agentive suffix but this is not foolproof, considering the example from (McGregor, 1993) where the VALA follows a noun but acts as a relativizer.

(10) टोपी **वाला** लड़का
hat **VALA** boy
the boy wearing a hat

It's very difficult to establish the PropBank frame for (10): wear-01 follows only from nuanced semantic understanding. It's consequently difficult to automatically produce AMR graphs for this relativizer sense of VALA with English notation, as there's no easy machine-inferable frame to join the 'hat' and 'boy' concepts. A possible solution is to consider the 'attribution' that VALA bestows to the noun that is relativized, and consequently take advantage of the **:domain** relation (see the Appendix for the AMR translation of this example)

Another use of the particle denotes a stative sense, the state of 'going to'. This stative sense of VALA can be captured by the **stable** aspect relation in AMR following (Donatelli et al., 2018). An example is in the Appendix.

(11)	जिस	दिन	वह	रवाना होने	**वाला**	था	उसने	सब	ठीक-ठाक	किया
	REL.OBL	day	he	leave be.GERUND	**going_to**	COP.PAST	he.ERG	everything	right.ECHOCOMPOUND	do.PERF

The gerund plus VALA combination is not a reliable heuristic to mark the stative, as the following examples show. These examples are similar to the 'teacher' example in English AMR (Banarescu et al., 2013) i.e. a person who teaches:

(12) नाचने **वाला** (dancer)
dance.GERUND **VALA**

(13) लड़ने **वाला** (fighter)
fight.GERUND **VALA**

To illustrate with one more particle 'भी' (BHI), meaning 'also':

(14) तब से   मैं   अजगर   के_अलावा   **कुछ_भी**   बनाना नहीं सीख पाया
    then ADP I.NOM python   other than   **anything** draw   NEG   learn <light verb>.PERF

The use of **कुछ** in (14) contrasts with its use in (20) below. In (20), **कुछ** does not come under the scope of any particle and keeps its original meaning of 'something'. In (14), it comes under the scope of the particle BHI and takes on the new meaning of 'anything'. In AMR, scope does not have a deep representation (Banarescu et al., 2019). The position of **भी** is represented with a **:mod** relation between the token and the particle.

BHI is also used for emphasis and intensification, as in (15) below. The closest translation could be to the adverb 'even' (McGregor, 1993). In such cases, the particle could be dropped from the AMR graph or it could be represented with the **:degree** relation as it is an intensifier.

(15) मैंने   अपना   राज्य   घूम   कर   देखा   **भी**   तो   नहीं है
    I.ERG REFL   kingdom   roam   CONJ see   **even** EMP   NEG   COP

5. Loss of morphological inflection

English AMR does not capture morphosyntax for tense nor aspect nor number (Banarescu et al., 2013) nor gender nor case. Hindi predicates and their arguments undergo morphological inflection for all five. All this information is consequently lost with English AMR notation, except for tense and aspect which are discussed below. Morphosyntax also plays a role in determining mood, modality, and in resolving zero anaphors or null subjects.

Using English AMR notation, special rules or heuristics become necessary to derive frames and concepts in the graph, based on the morphosyntax of the predicate and its arguments. These rules need to be applied by an annotator based on their knowledge of Hindi, or an automatic parser. Because English AMR doesn't represent morphosyntax in the graph, AMR to text generation would lose the style of the original sentence, and the mood and some cases of modality as well.

5.1 Tense and Aspect

Hindi has three tenses: past, present and future, and three types of aspect: habitual, progressive, and perfective (Koul, 2008). The extensions to AMR to capture tense and aspect (Donatelli et al., 2018) do capture interactions of Hindi tense and aspect in the indicative mood, which has separate morphology for the three tenses as well as separate morphology / secondary verb participles for the different aspect types. Tense can be captured using the **:time after / before / now** annotation conventions; progressive aspect can be captured using the **ongoing** convention, the perfective with the **ongoing / completed** convention, and the habitual with the **habitual** convention.

However, the presumptive mood, indicated exclusively by inflections on the primary copula होना (to be), does not include tense morphosyntax on the copula. Consequently, it's very difficult to pinpoint the tense referent to speech time, without a signal from any of the other lexical markers. This can be seen across aspect types as in the sentences below, taken from (Sharma, 2008). Some aspects filter out certain tenses, but still result in more than one tense to choose from. The PRESUMP gloss is used for presumptive mood below:

(16) रामू   [पिछले साल / आजकल / ~~अगले साल~~] बहुत आम   **खाता**   होगा
    Ramu   [last year / these days / ~~next year~~]   much   mango **eat.HAB**   PRESUMP

(17) रामू   [तब/ अब/ कल]   वेनिस   में   घूम   **रहा**   होगा
    Ramu   [then / now / tomorrow] Venice LOC   roam   **PROG** PRESUMP

(18) रामू   [कल / आज / ~~कल~~]   ज़रूर   वेनिस **गया**   होगा
    Ramu   [yesterday / today / ~~tomorrow~~] certainly Venice **go.PERF** PRESUMP

From an AMR context, tense will need to be determined from the explicit lexical signals in the sentence with the presumptive mood. Should the lexical signals be absent, the following options are available for the AMR: ignore tense and rely on the (new) **presumptive** attribute in the **:mode**, or add multiple, disjunctive **:time** relations to the AMR for all possible tenses that have not been filtered out by the aspect.

5.2 Modality and Mood

One way to signal modality is by a combination of dative case on the subject with an infinitive form of the main verb predicate, as in (19). These cases of modality could map to PropBank's obligate-01, but an extra rule or heuristic is need during parsing to determine the presence of that modal from the morphosyntax:

(19)	अब	**मुझे**	दूसरा	पेशा	**चुनना**	पड़ा
	now	**I.DAT**	second	profession	**choose.INF**	\<action towards self\>

Mood is captured using morphology on the verbs, light verbs, or primary copula. Hindi moods include the presumptive, subjunctive, imperative and indicative (Koul, 2008). The English AMR notation would need new attributes for the subjunctive and presumptive moods present in Hindi using the **:mode** attribute.

(20)	बस	बच्चे	खिड़की	से	नाक	रगड़ते	कुछ	देख	रहे	**होंगे**
	just	children	window	INST	nose	crush.HAB	something	see	PROG	**COP.PRESUMP**

5.3 Null subjects and zero anaphora

Some, but not all, sentences have null pronoun subjects or zero pronoun anaphors, which are resolved through the inflections on the predicate or copula, which agree with the pronoun in person, number, and gender:

(21)	अंदर	बैठे	सो	रहे	**होंगे**		या	जम्हाई_ले	रहे	**होंगे**
	Inside	sit.PART	sleep	PROG	**COP.PRESUMP.3ʳᵈ.PL**		or.CONJ	yawn	PROG	**COP.PRESUMP.3ʳᵈ.PL**

(21) has no explicit subject; the morphology on the primary copulas indicate that the subject is a third person plural pronoun, ये ('they'). In a framework such as MS-AMR (O'Gorman et al., 2018) or DOC-AMR (Naseem et al., 2021), a heuristic is needed to determine the presence of zero-pronoun anaphors or null subjects, resolve them from the morphology and then find their antecedents, if existing, to form the required coreference links. The AMR parser or annotator would need to introduce either morphosyntax or derived pronoun concepts into the graph.

Resolving zero pronoun anaphora is not always straightforward. A complication comes from the use of the honorific address 'आप' which is used for both singular and plural subjects (Koul, 2008). The honorific always takes plural third person morphology, even if the antecedent is singular and even when the context calls for a second-person subject:

(22)	एक	क्षण	को	तारे	के_नीचे	इंतजार	**करिएगा**
	one	moment	ADP	stars	under	wait	**3ʳᵈ.PL.IMP**

Only the imperative mood in the subjectless (22) gives a clue that the zero anaphor resolves to the second-person honorific (The second person honorific antecedent here does exist in previous sentences in the corpus). This clue is not always present:

(23)	उसे	कुकुरमुता	**कहेंगे**
	he.OBL	mushroom	**3rd.FUT**

(23) is in indicative mood. The context from the previous sentence in the Hindi Little Price, and the lack of an antecedent in previous sentences in that corpus indicate that the null subject is a generic 'they' or 'people'. However, the morphology allows for the possibility of the null second person honorific subject, and it's difficult to resolve the null pronoun subject given this choice using morphosyntax from the sentence alone.

6. Appendix

Detailing desiderata:

6.1 AMR Annotations

AMR annotations for the subset of the Hindi version of The Little Prince are present in github:
https://github.com/nitinvwaran/littlepricehindi-amr.git
The repository also contains the subset of AMRs from the English Little Price for the selected Hindi sentences.

Selected AMR annotations are reproduced below:

Example 9:
राजकुमार को अपने कर्तव्य के प्रति इतना ईमानदार बत्ती वाला बड़ा भला लगा
(f / लगा.feel-01
 :ARG0 (p / राजकुमार.prince)
 :ARG1 (g / भला.good
   :degree (v / बड़ा.hugely))
 :ARG2 (l / बत्तीवाला.lampman
  :ARG1-of (f2 / ईमानदार.faithful-00
   :ARG2 (d / कर्तव्य.duty
    :poss अपने.l))))

Example 10:
टोपी वाला लड़का
(t / टोपी वाला.hat VALA
 :domain (b / लड़का.boy))

Example 11:
जिस दिन वह रवाना होने वाला था उसने सब ठीक-ठाक किया
(r / ठीक-ठाक किया.right-01
 :ARG0 (h / उसने.he)
 :ARG1 (e / सब.everything)
 :time (d2 / दिन.day
  :mod (d / रवाना होने.depart-01
   :ARG0 h
   :time (a / before :op1 (n/now))
   :stable -)))

Example 14:
तब से मैं अजगर के अलावा कुछ भी बनाना नहीं सीख पाया
(l / सीख.learn-01
  :polarity -
  :ARG0 (i / मैं.i)
  :ARG1 (m / बनाना.draw-01
  :ARG0 i
  :ARG1 (k / कुछ
  :mod (b / भी)))
 :time (f / से.from
  :op1 (t / तब.then))
 :ARG2-of (e / के_अलावा.except-01
  :ARG1 (p / अजगर.boa_constrictor)))

Example 15:

मैंने अपना राज्य घूम कर देखा भी तो नहीं है

(s / देखा.see-01 :polarity नहीं.-
   :degree (b / भी)
   :ARG0 (i / मैं.i)
   :ARG1 (k / राज्य.kingdom
     :poss मैं.i)
   :time (a / कर.after
     :op1 (s2 / घूम.stroll-01
       :ARG0 i
       :ARG1 k)))

Example 19:

अब मुझे दूसरा पेशा चुनना पड़ा

(o / obligate-01
   :ARG1 (m4 / मुझे.me)
   :ARG2 (c4 / चुनना.choose-01
     :ARG0 m4
     :ARG1 (p3 / पेशा.profession
       :ord (o2 / दूसरा.ordinal-entity :value 2)))
   :time (n2 / अब.now))

Example 20:

बस बच्चे खिड़की से नाक रगड़ते कुछ देख रहे होंगे।"

(a2 / and
   :op1 (c / रगड़ते.crush-01
     :ARG0 (c2 / बच्चे.child
       :mod (o / बस.only))
     :ARG1 (n / नाक.nose
      :part-of c2)
     :ARG2 (w / खिड़की.window)
     :habitual +)
   :op2 (s2 / देख.see-01
     :mode presumptive
     :ongoing +
     :ARG1 (s3 / कुछ.something)))

Example 21:

अंदर बैठे सो रहे होंगे या जम्हाई ले रहे होंगे,

(o / या.or
   :op1 (s / सो.sleep-01
       :mode presumptive
       :ongoing +
       :ARG0 (t2 / they)
       :location (i2 / अंदर.inside))
   :op2 (y / जम्हाई ले.yawn-01
       :mode presumptive
       :ongoing +
       :ARG0 t2
       :location i2))

Example 22:

एक क्षण को तारे के नीचे इंतजार करिएगा

```
(w / इंतजार करिएगा.wait-01
    :mode imperative
    :ARG1 (y / HONORIFIC you)
    :ARGM-TMP (m / क्षण.moment :quant एक.1)
    :location (u / के नीचे.under
        :op1 (s / तारे.stars)))
```

6.2   Light Verbs

A table of light verbs in Hindi and their meaning follows from (Koul, 2008). The second column is the traditional meaning of the verb, which is lost when used in a compound / conjunct verb.

आ a:	come	change of state from within
जा ja:	go	change of state
ले le	take	action for or toward others
पड़ paṛ	fall	action for or towards self
दे de	give	change of state, suddenness
जा ja:	go	direction away, simple completion
डाल ḍa:l	throw	speed, recklessness, relief, completion
छोड़ choṛ	release	psychological separation, relief
रख rakh	put/keep	proactiveness, future use in view
बैठ bɛṭh	sit	action for or towards self
उठ uṭh	rise	action for or towards self
पहुंच pahũc	reach	action for completion, direction
चल cal	walk	direction away, completion
मर mar	die	completion, lack of control
मार ma:r	kill	change of state, suddenness

6.3   Linguistic Glosses

Glosses are taken from: https://en.wiktionary.org/wiki/Appendix:List_of_glossing_abbreviations
Where the gloss is new, it has been explicated.

7.   References

Arora, A., Venkateswaran, N., & Schneider, N. (2021). SNACS Annotation of Case Markers and Adpositions in Hindi. SCiL.

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). Abstract Meaning Representation for Sembanking. *The Association for Computer Linguistics, In S. Dipper, M. Liakata&A. Pareja-Lora (eds.), LAW@ACL*, 178–186.

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2019). *Abstract Meaning Representation (AMR) 1.2.6 Specification*. https://github.com/amrisi/amr-guidelines/blob/master/amr.md

Caturvedi, M. (1970). *A practical Hindi-English dictionary*. National Publishing House.

Donatelli, L., Regan, M., Croft, W., & Schneider, N. (2018). Annotation of Tense and Aspect Semantics for Sentential AMR. *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions*, 96–108.

Kingsbury, P., & Palmer, M. (2002). From Treebank to PropBank. 2002. In Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002), Las Palmas, Spain. *Proceedings of the 3rd International Conference on Language Resources and Evaluation*.

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

McGregor, R. S. (1993). *The Oxford Hindi-English dictionary*. Oxford University Press.

Naseem, T., Blodgett, A., Kumaravel, S., O'Gorman, T., Lee, Y.-S., Flanigan, J., Astudillo, R. F., Florian, R., Roukos, S., & Schneider, N. (2021). *DocAMR: Multi-Sentence AMR Representation and Evaluation*. arXiv:2112.08513

O'Gorman, T., Regan, M., Griffitt, K., Hermjakob, U., Knight, K., & Palmer, M. (2018). AMR Beyond the Sentence: The Multi-sentence AMR corpus. *Association for Computational Linguistics*, *Proceedings of the 27th International Conference on Computational Linguistics*, 3693–3702.

Sharma, G. (2008). A Pragmatic Account of the Hindi Presumptive. *Annual Review of South Asian Languages and Linguistics 2008, Mouton de Gruyter*, 83–116. https://doi.org/10.1515/9783110211504.1.83