

INGIT: Limited Domain Formulaic Translation

from Hindi Strings to Indian Sign Language

Purushottam Kar, Madhusudan Reddy, Amitabha Mukerjee and Achla M. Raina

Indian Institute of Technology Kanpur

Kanpur -208016, India

{purushot, msreddy, amit, achla}@iitk.ac.in

Abstract

We report a cross-modal translation system from Hindi strings to Indian Sign Language (ISL) for possible use in the Indian Railways reservation counters. INGIT adopts a semantically mediated formulaic framework for Hindi-ISL mapping. An in-depth investigation into the structure of ISL forms the groundwork for INGIT. Some representational and mapping issues concerning cross-modal translation are identified and an implementation design is evolved. We adopt the Construction Grammar approach for handling formulaic inputs in terms of a construction lexicon with single constituents as well as larger phrases, with direct semantic mappings at each level. We present results based on a small corpus collected at a railway counter, for which translations were validated from native ISL signers. The work builds upon a semantic module worked out for Hindi and ISL.

1 Introduction

One out of five deaf people in the world live in India, yet the Indian deaf community is singularly disenfranchised owing to isolation in the society at large and the oralist tradition prevalent in deaf schools (Deshmukh, 1996). Indian Sign Language (henceforth ISL) is a sign language variety used in India (Zeshan 2000).

Here we present a prototype system designed as a proof-of-concept for a Hindi to ISL translator in the railway reservation domain, a common public need for citizens. The system,

named INGIT¹ translates input from the reservation clerk into Indian Sign Language, which can then be displayed to the ISL user. INGIT currently accepts transcribed spoken language strings as input and generates ISL-gloss strings which are converted to a graphical display via HamNoSys (Prillwitz et al 1989) simulation. Only the utterances of the reservation clerk are translated since the deaf client can respond via the paper form.

Most translation systems decompose the input into separate syntactic and semantic modules; in contrast, INGIT adopts a formulaic approach (Wray *et al.*, 2004). Here both syntactic and the semantic mappings are stored in a *constructicon*, which lists larger constructions along with single constituents.

The objective in the present work is to create a scalable system which would be fully developed based on a much larger interaction corpus (the present corpus of 230 sentences and video-translations was extremely small). Also, while the corpus was validated by signers, we could not actually record any sign language transactions at a railway counter, since no such counters exist. Clearly, for a larger corpus, some design decisions may change, and our objective in encoding formulaic constructs was that the formulaic nature of the multi-word constructions were more likely to be retained than single constituents. Overall coverage based on compositional approaches were minimally deployed since these rules are often subject to considerable tweaking as the data changes. Also

¹ INGIT is a Sanskrit word meaning signed, and has the connotation of a gestural sign. It was once hoped that it might stand for *INDian Gestural Interaction Translator* but this expression was unwieldy and now it is just an unexpanded name.

the objective was to create a template that would be amenable to the development of other similar public domain interaction systems.

Since Indian Sign Languages are yet to be analyzed in much detail, one of the challenges was the characterization of the fragment of ISL that arises in such transactions and the cross-modal issues in going from speech to sign.

1.1 Sign Language Modality

ISL is a spatial language and one consequence of this modality is that it has at its disposal multiple channels of communication (hand, body, face), resulting in parallel communication streams, thus differing sharply with the linear nature of oral languages. Also, there is a greater degree of iconicity in its symbols, and simultaneous events may be signed as such - e.g. “while the teacher is teaching, you should observe”, would use one hand for signing *teacher*, and simultaneously the other signing *observe*. The other side of this parallelism is that by and large sign production is half the speed of oral production (Sexton, 1999); clearly this is compensated for by various means such as non-manual signs and elision of semantically non-salient constituents.

As in other sign languages, ISL uses non-manual signs (primarily facial) in parallel with manual signs (hand/arm) to indicate negations questions, and suggestive phrases. INGIT handles this parallel aspect— facial expressions are specified for negations and interrogatives and suggestive phrases, and the scope of these expressions are demarcated.

Another difference in Sign is that spatial deixis is used for directional verbs and anaphora. This latter is a great challenge for oral-to-Sign translation since anaphora and other reference mechanisms are handled by indicating the previous spatial position where this entity appeared (spatial deixis), often abbreviating the sign for the object by using a classifier. Thus the system has to explicitly identify the anaphora referent, and store the location where each referent appears. Unlike in many oral-to-oral translation systems, anaphora cannot be passed intact from the source onto the target language, a significant hurdle to non-discourse models. Our system is a one-way interaction, and many spatial referents are lost – in a normal discourse, the system should have observed the location and manner of the sign articulated by the speaker in the first place. What an ISL speaker would have done in this position is to device a default deixis,

and this is what INGIT does. Thus referents such as “that train” are passed on as TRAIN-DEI which were easily handled by our Sign interlocutors in context. For more general oral-to-Sign systems however, this is an aspect that would need to be handled.

In addition, the input text in situations involving close limited interactions like at the railway counter may elide many arguments. INGIT handles problems such as elliptic omissions using simple default assumptions about the domain. Several commonly found idiomatic structures of Hindi were also handled easily in the formulaic structure.

2 Existing Research

Even in relatively well-developed Sign research communities such as ASL (US), BSL (UK) or Japanese Sign, research on cross-modal translation is sparse. As in the MT community in general, one may characterize the work roughly into a) *Form*-based, where surface forms are mapped using certain systematic alternations (Veale and Conway, 1994; Lemcke 1997, Grieve-Smith 1999, Zhao *et al*, 2002, Speers, 2002), and b) *Semantic*-based, which attempt a more semantic approach (Marshall and Sáfár, 2003, Wray and Cox 2004).

Form-driven approaches include ZARDOZ by Veale and Conway (1994) which processes English input serially using morphological analysis, idiomatic reduction, and parsing by a unification grammar. Metonymic and metaphoric references are removed, based on which BSL (in HamNoSys notation) is generated. Lemcke (1997) builds a translator for ASL but does not handle non-manual signs. Grieve-Smith (1999) uses a syntactic approach for generating ASL translations in the weather domain, but the Sign production is very refined. A surface mapping based on TAG grammars is used in (Zhao *et al.*, 2000), who also support topicalized orderings. Inflectional aspects are mapped onto Sign via parameters like speed and force which result in morphological variations. Negation and interrogatives are handled using non manual signs. Speers (2002) uses correspondence rules to map English f-structure (syntax) into ASL, where it is used to generate the phrase-structure. There is a thorough analysis of Sign production and the effect of different phonotactic environments.

Semantically motivated models can pursue traditional methods, where syntactic

combinations are given semantic interpretations (Marshall and Sáfár, 2003), or a formulaic or construction based approach, where direct mappings may exist for larger units (Wray and Cox 2004). Marshall and Sáfár (2003) model the discourse via a Discourse Representation Structure (DRS) to handle anaphora and other discourse phenomena. The oral DRS is converted into a Sign (BSL) DRS from which an equivalent HPSG semantic structure is converted into HamNoSys.

In translation based on compositional semantics, the parsed structures retain aspects that do not transition from a spoken mode to the Sign mode, and these need to be rectified via schematization which can be of considerable complexity and are difficult to maintain. Also, as in any serial translation system, the overall accuracy requires the accuracy of each stage to be high, and interdependence between stages makes it difficult to tune each stage separately. Further, since most of the grammars for spoken languages do not handle parallelism in scope and topic identification, so where these aspects will fit into the mapping is often unclear.

These difficulties are to some extent overcome in formulaic or construction-based approaches, where larger strings occurring frequently are accorded unit status, and only structures without a direct construction would be handled compositionally. Wray *et al.* (2004), in their TESSA system, present a Sign translation system based on a purely formulaic approach (with minimal compositionality). Here the input spoken expression is mapped to one of several predefined “target phrases” (all paraphrases of a target string generate the same translation). The speech recognition phase itself attempts this mapping after which the whole input expression can be analyzed as composed of either concatenated target phrases and/or target phrases having open slots in them to be filled by numerals/dates and the like. TESSA opts for semantically-based translation using a probabilistic framework to express the given message. While this work addresses the issue of directly handling semantic mappings with larger constructions, it can only produce those expressions it is designed for, and no other inputs can be handled. Also parallelism involving non-manual signs does not appear to be handled.

While not moving so completely towards formulaic structures, INGIT uses constructions to encode larger units in the input.

2.1 Construction Grammars

We use Construction Grammars (Kay and Fillmore, 2001) as our vehicle for implementing a formulaic grammar. Construction Grammars commit themselves to the parity of linguistic expressions irrespective of their structural complexity. Compositionality of expressions in natural language is a matter of degree. We encounter completely opaque utterances like

i) ... fly in the ointment ...

relatively less opaque ones like

ii) What is this fly doing in my soup?

which display a template like structure as well as compositional ones like

iii) The fly is buzzing.

Construction Grammars treat general patterns in the language which account for the compositional utterances and the more idiomatic ones equally. To do so constructions can be proposed at various levels, viz. morphological, lexical, phrasal, sentential. Constructions are essentially form-meaning mappings. These mappings are bidirectional and can be used for production as well as parsing. The only operation defined in the grammar is that of unification in which constructions can be unified to form higher level constructions subject to constraints.

The constructions are stored in what is known as a constructicon as there does not exist a distinction between the lexicon and the grammar. A consequence of constructions being form-meaning mappings is that the parsing process does not involve generation of a parse structure from which the semantics can be inferred. Instead the semantic structure itself is a natural outcome of the parsing process. Thus general patterns such as word order as well as the more idiomatic expressions are handled at the same level.

INGIT uses a hybrid-formulaic approach with many larger formulaic units, along with composition involving single constituents where larger constructions are not found in the input. Constructs such as negation and interrogatives are handled through parallel non-manual generative devices. We use Fluid Construction Grammar (FCG) (Steels and Beule, 2006), a computational model that encodes paired syntactic and semantic structures or constructions. The rules so specified in terms of the constructions are bidirectional and hence used for both parsing as well as generation for which a unification based approach is adopted. FCG permits additional structures which are

purely syntactic or purely semantic to be proposed to allow syntactic/semantic processing.

The output of the FCG-based ISL construction is a set of strings (ISL-gloss) which is passed to a rudimentary graphics engine that is designed to accept ISL strings tagged with non-manual markers (Section 4.8).

2.2 INGIT : Architecture

INGIT works on strings of transcribed Hindi spoken text. A domain-specific construction grammar for Hindi, implemented in FCG, converts the input into a *thin* semantic structure which is input to ellipsis resolution, after which we obtain a *saturated* semantic structure. Depending on the type of utterance (statement, query, negation, etc) a suitable ISL-tag structure is generated by the ISL generator. This is then passed to a HamNoSys converter to generate the graphical simulation (Figure 1).

For validating the system, a small corpus was collected on six different days, based on interaction with speaking clients at a computer reservation counter, and constituted 230 utterances, of which many were repeated. The vocabulary of 90 words included 10 verbs in various morphological forms, 9 words related to time, 12 words specific to the domain (e.g. ticket, tatkaI, etc.), pronominal words including anaphoric referents, question words and function words. Other words were numerals (15), names of months (12), cities (4) and trains (4) as well as digits particles etc.

To get started with our work, we took this corpus, had various phrases converted in different ways by ISL interlocutors, and started analyzing the resulting Sign strings to come up with the ISL construction. In the next section, we present some of the ISL mappings that result from these sentences.²

3 Structure of Cross-Modal Mapping

One of the challenges in working with ISL is the insufficient characterization of the language itself. At the outset, we characterize some of the corresponding structures in ISL in view of the cross-modal mapping to be performed. Consider the following examples:

- 1) शताब्दी कानपुर नहीं जाती है
 shatAbdi kAnpur nahiN jAti hai
 => {SHATABDI @n{KANPUR GO NEG}}

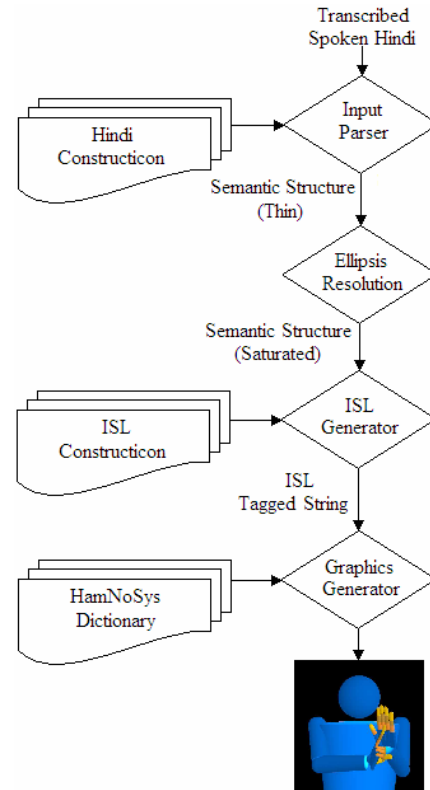


Figure 1. Architecture of the INGIT System. Here “thin” semantics implies that some arguments may be elided. These are filled in by the ellipsis resolution module, resulting in a fuller Semantics which is used to generate ISL.

The last line is the ISL-gloss, which is a form of written Sign where symbols such as SHATABDI and GO are tokens for ISL signs. Later, during production, these would be instantiated based on the HamNoSys dictionary. The @n tag indicates a parallel non-manual instance of negation, the scope for which is indicated using parentheses. Visually, this reflects a facial expression persisting during the signing of the negated phrase (Figure 2). In ISL it is often reinforced at the end by a manual NEG (as in this example).

- 2) राजधानी रात में चलती है
 rAjdhani rAt mein chalti hai
 => {RAJDHANI NIGHT GO}

We observe that the particle में (*mein*, at) which serves a grammatical function in Hindi without a counterpart in ISL. This is handled using a compositional construction.

² see <http://www.cse.iitk.ac.in/users/language/sign> for the video-tagged data.

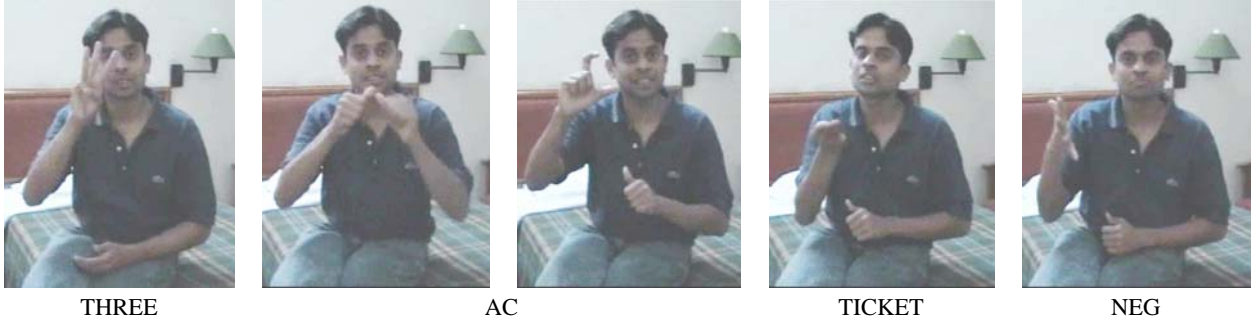


Figure 2: An ISL Signer signing the string 3-ए.सी में टिकट नहीं है / 3-A.C. meiN ticket nahiN hai
=> {THREE AC TICKET NEG}. Note the facial expression in the last sign

3) दस रुपये दीजिए

das rupaye dijiye
=> {MONEY TEN GIVE {YOU ME}}

This illustrates an elliptic omission which could be mapped onto other spoken languages without deficit, but in Sign, the participants need to be expressly demarcated (See Section 3.1). This requires ellipsis resolution and it is handled as a formulaic construction at the parsing stage (See Section 4.3).

4) टिकट नहीं मिलेगा क्योंकि वेटिंग है

ticket nahiN milegA kyonki
waiting hai
=> {{TICKET @n{GET NEG}} @q{Q-WHY}
{WAITING-LIST}}

Here two individual constructions are hierarchically combined using the क्योंकि (*kyonki*, because) construct. The @q construct indicates a non-manual interrogative the scope for which is indicated by the parenthesized string. The वेटिंग है (*waiting hai*, is-waitlisted) input is handled formulaically using the construction (6) below.

5) आप राजधानी ले लीजिए

Ap rAjdhAni le lijiye
=> {RAJDHANI @c{YOU}}

The @c construct reflects non-manual suggestion or counsel usage.

6) x में y वेटिंग है

x mein y waiting hai
=> {x WAITING-LIST y}

This is a simple formulaic construction which takes two other constructions x and y and generates the appropriate output.

7) x में वेटिंग है

x mein waiting hai
=> {x WAITING-LIST}

This is same as (6) except that y is dropped - reflecting a limitation of construction grammars like FCG – it is difficult to specify optional arguments.

Based on our corpus data, we find that the cross-modal mappings can be handled either at the constituent level (compositional), or are mapped as a unit (formulaic). These are described next.

3.1 Constituent Level Mappings

Composition involving single constituents were observed to involve either a complete correspondence as in (1), exhibiting only constituent reordering, or a partial map which could involve constituent deletion (2) or constituent insertion (3).

We observe that constituent deletion (barring cases of ellipsis) involves omission of functional constituents like the temporal post position में (*mein*, at) in (2).

Also, constituent insertion, as in (3), involves an explication of constituents elided in the spoken expression. As mentioned earlier, Sign requires referents to be specified spatially, and one does not have the freedom of passing instances of ellipsis onto the target language. Argument roles for predicates of dyadic or triadic type are specified using directionality of the sign for the predicate. In (3), the arguments YOU and ME are trivially located in space. The argument roles of the donor and the recipient are explicated using the direction of the sign for GIVE which directs from the donor to the recipient. Thus, ellipsis resolution must be performed for generating a correct translation.

3.2 Formulaic (Unit) Mappings

In many cases, mappings involve major shifts in the constituent set, or the expressions were found to describe a frequent pattern, signifying unit usage. These included compositional constructions as in (4) where we observe a mapping from an affirmative reason clause to a content (why) question. (5) merits a construction level treatment since the expression has a

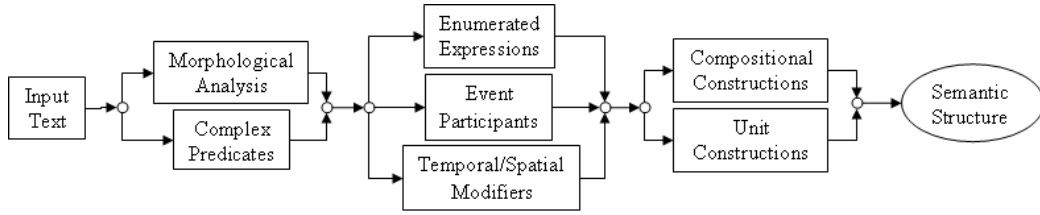


Figure 3. Overview of the Input Parser

suggestive mood which is expressed through non-manual markers in ISL. This mood is captured holistically by the use of slotted templates.

3.3 Anaphoric Expressions

Anaphora resolution through discourse analysis is a task commonly performed by cross-modal translation systems (e.g. Marshall and Sáfár, 2003). However given our one-way discourse, many spatial referents are missing, and we found it adequate to use default deictic references. Consider the following example:

8) वह गाडी कानपुर नहीं जाएगी
 wah gADi kAnpur nahin jAYegi
 => {TRAIN -DEI @n{KANPUR GO NEG}}

Here the deictic sign signaled by -DEI is contextually deduced by our ISL listeners to indicate the particular train in question even though the spatial position indicated by this deictic sign is not a spatial node that was previously defined for that particular train. Thus, while discourse elements have not been implemented in INGIT so far, it may be possible to go some distance (in this limited domain) without invoking that heavy machinery.

3.4 Polysemous Expressions

Polysemous expressions pose problems for all translation systems. The following examples put the problem in the context of INGIT:

9) राजधानी में दस वेटिंग है
 rAjdhAni mein das waiting hai
 => {RAJDHANI WAIT-LIST TEN}
 10) आपके पास फार्म है
 Apke pAs form hai
 => @q{YOU-HONORIFIC FORM IS-EXISTIVE}

Clearly the lexical item है (*hai*, be) in (9) is exhibiting attributive character with the sense of the word वेटिंग (*waiting*, waiting) actually being wait-listed. However in (10), है (*hai*, be) describes the existence of a possession in an alienable sense. ISL recognizes these multiple senses of है (*hai*, be) to be distinct and expresses

them differently. The following section will describe the system architecture of INGIT and will present solutions to the various problems posed above.

4 INGIT System Details

Based on the above analysis, INGIT adopts a formulaic approach that directly generates the semantic structure where possible (about 60% cases), and defaults to a compositional mode for the others. The main modules in the system are:

- Input Parser
- Ellipsis Resolution Module
- ISL Generator (including ISL lexicon with HamNoSys phonetic descriptions)

4.1 Extending the FCG Framework

INGIT uses the FCG framework both for analyzing the input (oral) and for generating an output (Sign). The issues related to elliptical expressions motivate a semantically mediated approach towards translation process as ellipsis resolution is not possible unless the event structure is accessible. Every expression is analyzed with respect to its syntactic and semantic structure in the parsing as well as generation stages. The FCG engine was extended in this instance by the ellipsis resolution module, which is implemented directly in LISP and functions as an intermediary between the Input Parser and the ISL Generator.

4.2 Input Parser

INGIT accepts as input transcribed spoken language strings which may be tagged for intonation patterns. Currently the system handles only one such intonation tag which was frequently observed in our corpus. This is the '?' tag for affirmative questions which often occur without a question word. Consider the following input string:

11) कल जाना है?
 kal jAnA hai ?

“(Do you) want to go tomorrow?”

4.3 The Translation Process

We now consider details of the translation implementation. Consider the sentence,

12) शताब्दी शाम को कानपुर नहीं जाती

है

shatAbdi shAm ko kAnpur nahiN
jAti hai

“Shatabdi does not go to Kanpur in the evening.”

First, the verb-auxiliary complex जाती है (*jAti hai*, goes) is morphologically analyzed and its root is identified as जा (*jA*, go). The semantic structure for जा (*jA*, go) appearing in the constrict-icon, is as follows:

((MOTION-VERB EV) (GO EV)
(VERB-CLASS EV UNARY)
(ARGUMENT-1 EV OBJ)
(TIME-FRAME EV X)
(ARGUMENT-1-PREREQ EV MOBILE))

which merely states that it is a verb in the motion class, and it takes an object as its single obligatory argument.

Next, शाम को (*sham ko*, in the evening) is recognized as a temporal modifier and कानपुर (*kAnpur*, kanpur) gets identified as a spatial modifier.

In the subsequent step, the word order of this expression is seen to be matching that of the compositional construction {SUB-NOMINATIVE MODIFIER-1 MODIFIER-2 NEGATION UNARY-VERB}. Thus the valence items in the semantic frame are identified as the lexical items शताब्दी (*shatAbdi*, shatabdi) whereas शाम को (*sham ko*, in the evening) and कानपुर (*kAnpur*, kanpur) serve as optional time and destination modifiers. The negation operator नहीं (*nahiN*, not) is identified and its scope is marked as the negation of the corresponding VP and the corresponding semantic structure is generated from the FCG engine:

((MOTION-VERB X-95) (GO X-95)
(VERB-CLASS X-95 UNARY)
(ARGUMENT-1 X-95 X-96)
(TIME-FRAME X-95 PRESENT)
(SHATABDI X-96)
(DISCOURSE-ROLE X-96 EXTERNAL)
(GENDER X-96 FEMININE)
(MOBILITY X-96 MOBILE)
(NEG X-73)
(KANPUR X-61)
(EVENING X-58)

(TEMPORAL-MODIFIER X-30 X-58)
(SATURATED X-41) (EVENT X-41 X-95)
(MODIFIERS X-41 X-95 X-30 X-61)
(NEGATOR X-41 X-95 X-73))

Here the X-*nn* are semantic referents – e.g. X-95 is a motion verb, specifically GO, which reflects a unary predicate with argument X-96 (शताब्दी). Similarly शाम को is identified as a temporal-modifier. Thus the event X-95 has X-30 (शाम को) and कानपुर as modifiers. The entire event X-95 is negated by the referent X-73.

The above demonstrates a compositional process for input parsing. For inputs that match a unit construction (or phrases participating in composition that match such a construction) the direct semantics map for the input will be immediately generated, or if a sub-phrase, it will be passed to the appropriate structure.

If the semantic structure is saturated, it is passed directly to the ISL generator; else it is passed to the ellipsis resolution module which attempts to fill in any elided arguments.

4.4 Ellipsis Resolution Module

Consider the sentence

13) दस रुपये दीजिए

das rupaye dijiye

“Give ten rupees”

Our procedure identifies the verb *give* and the object दस रुपये (*das rupaye*, ten rupees) as an enumerated expression. However, the predicate दीजिए (*dIjiye*, give) in the above expression takes three arguments according to the constructicon, whereas the expression provides only one. Thus the semantic structure generated would be {OBJ-ACCUSATIVE VERB-TERNARY}, which is clearly incomplete and thus the construction is not saturated.

We observe that all discourse in our corpus reflects a two-participant constraint (speaker and the listener) – i.e. elided constituents are found within these two. Based on this, INGIT currently handles elision of participants in ternary events and those of the subject in unary and binary events. Here, the morphology of दीजिए (*dIjiye*, give-2nd-pers-honorific) indicates that and its donor be the addressee and that both participants be animate beings. The donor is thus identified as YOU. Since the intended recipient cannot be the same person as the donor in this type of utterance, the remaining animate being, i.e. the speaker, becomes the recipient thus saturating the semantics of the event.



Figure 4. *HamNoSys Notation for शताब्दी शाम को कानपुर जाती है*

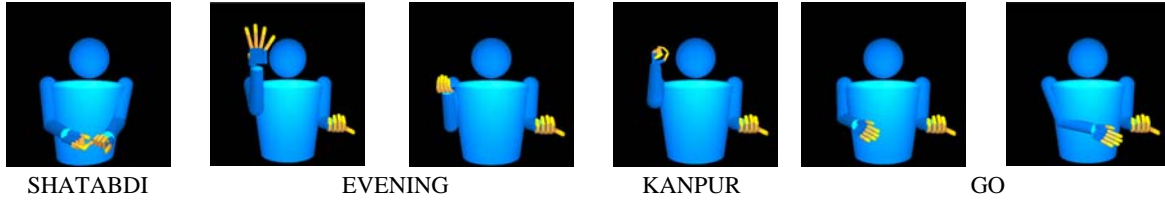


Figure 5. *Graphic Simulator Output for शताब्दी शाम को कानपुर जाती है*. The symbols above each token constitute its HamNoSys transcription from the HamNoSys lexicon.

4.5 ISL Generator

ISL forms for word roots are mapped from the semantic tokens subject to further morphological inflections. These ISL-tag tokens are handled bottom-up – i.e. modifiers and other smaller units form lexical groupings in ISL-specific word order. Next ISL-constructions specify features like word order and scope of non manual tags.

For example in the sentence शताब्दी शाम को कानपुर नहीं जाती है, after word roots like SHATABDI, EVENING and GO have been identified, the ISL template which finds the semantics matching its own i.e. {SUB @n{MODIFIER-1 MODIFIER-2 UNARY-VERB NEGATION}}. Negation and Q scope resolution and constituent reordering takes place at this stage generating the word order and the output {SHATABDI @n{EVENING KANPUR GO NEG}}.

4.6 Coverage

More than half of the sentences in our corpus were repetitions or close paraphrases, constituting only 20% of the unique utterances. These were modeled as formulaic, in view of their frequency. Of the remaining, a small fraction were considered idiomatic and unsuited for a compositional approach (about 5%). The rest were currently being considered compositionally. As stated earlier, our proof-of-design construction was focused more towards handling the formulaic inputs, which are likely to undergo fewer changes as the system evolves. Thus the small construction reported here consists of 9 constructions for detecting smaller lexical groups, 15 top level compositional constructions and 8 top level unit constructions – a total of 32 constructions.

Coverage was not an important focus at this point, since any decisions based on such a small

corpus would no doubt be subject to change as more data arrive. The minimal compositional lexicon (as in many small hand-crafted grammars) failed on 23% of the input. Most of these failure cases would be relatively simple to account for through additional rules, but may interfere with other unseen utterances, or may be encoded as part of a formulaic approach, so we chose to wait before making decisions on compositional constructions. Here are some examples that are currently not handled:

- 14) अभी कोई गाड़ी नहीं मिलेगी
 abhi koi gadi nahin milegi
 => {NOW @n{TRAIN NEG} FULL}
- 15) पूरे जून में खाली नहीं है
 poore june mein khali nahin hai
 => {JUNE ONE_MONTH @n{TICKET NEG} FULL}

Of these, while (14) would be handled easily enough, (15) may be a little more complex owing to the elided elements. On the other hand, some structures as in (16) would require discourse level analysis which has not been handled in this system. This exhibits an interesting cross-modal disparity:

- 16) 6 जून तक है
 => {JUNE 6 UPTO IS-EXISTIVE}

Here we find that native Sign speakers do not accept the elided equivalent: {JUNE 6 UPTO IS-EXISTIVE} as a valid utterance, preferring instead to include the missing referent: {TICKET JUNE 6 UPTO IS-EXISTIVE}. Thus this type of structure cannot be handled without more extensive discourse referents.

4.7 HamNoSys Notation

At this point, we have the ISL gloss, which is now passed to the ISL generator for Sign production. Each token is converted into

HamNoSys (Prillwitz et al., 1989) which is a Sign notation system widely used to write Signs.

Each sign in HamNoSys is modeled by specifying parameters related to Hand Configuration, Hand Orientation, Palm Orientation and Hand Location. Further specifications describe motion, hand symmetry and a few other aspects.

Thus the sign EVENING has the HamNoSys representation:



Here the open-hand symbol specifies the hand-shape followed by a caret for the upward hand-orientation. The next two symbols specify the palm orientation and the hand-location (close to the head. This is followed by three signs indicating a slight downward motion as viewed from the front; a change in configuration during the motion; and the final hand configuration (fingers converging to a point). For each of these signs we have constructed graphical simulation modules for instantiating them.

Clearly this is a very constraining assumption, since in Sign as in any other system, production is much more than word (or phoneme) concatenation, and due to practical considerations, our approach is based on very coarse phonological granularity. The result is that the output is not very fluid and natural.

Some other aspects of Sign generation of broader interest were not handled in the present work. These include *assimilation*, e.g. in ASL, the sign for "me" may be combined with a sign such as "Indian" to indicate "I am Indian", and *geminatio* – if the final hand-pose in a segment is similar to the first pose in the following segment, it exhibits a sustained hold (Speers, 2002). While none of these features would be needed in scaling up INGIT within the railway counter interaction domain, they are essential for fluid Sign production in more general translation scenarios.

4.8 Graphical Simulator

Several approaches to Graphic generation are available including virtual-human based models from standard SL notations like HamNoSys (Marshall and Sáfár, 2003; Banerjee and Mukerjee, 2005). Other graphical simulations are based on MPEG-4 human models (Papadogiorgaki et al., 2004). In our approach, we convert our output ISL-tag strings to HamNoSys and model the output as a sequence of Signs. Non-manual tags such as @n are

reflected in a facial expression that persists during the scope of the negation Sign. The deictic marker used -DEI maps to an indexical gesture to an unallocated region in space. Transactional verbs e.g. दीजिए (*dljiye*, give) requires argument roles to be specified through directionality for which the arguments (in our 2-person discourse) are usually located trivially i.e. YOU and ME.

The graphical simulator converts the ISL-tags into HamNoSys (Figure 3) and displays these on a graphic simulator (Figure 4). Currently the graphical system's support for facial expressions is not complete and these are not shown.

5 Conclusion

The reported work focuses on the problem of cross-modal translation arguing for a semantically mediated translation procedure for cross-modal translation systems. A hybrid formulaic system is proposed. A working implementation for a small domain corpus of interactions from a railway booking counter is used to test the system.

The current system is clearly preliminary, and can only be validated by a much larger interaction corpus than was used here. A larger database would also permit a more systematic design of the construction.

Like all translation systems, this system also faces limitations on account of not being able to capture to pragmatics in certain situations. For example the sentence

17) अभी टिकट नहीं मिलेगा
 abhi ticket nahiN milegA
 => {NOW @n{TICKET GET NEG}}

This translation, though not completely wrong, would score low on native user acceptability, who would prefer {NOW @n{TICKET NEG} FULL}. While one can attempt ad hoc solutions for these situations based on unit constructions, this is clearly not a desirable approach for scalability. Such pragmatic considerations will remain a challenge for translation systems possibly until we have mechanisms that can learn semantic mappings from grounded interactions.

In the next phase of this project, we will be significantly extending our corpus and the corresponding video database of ISL sentences. Also, one of the immediate goals is to record sign-user interactions at a railway reservation counter, to observe if there are any differences arising from the mode of transaction.

Also, support has to be built so that the system can take speech as an input, which is of course a much larger issue. Another pressing need is to be able to describe ISL in terms of a framework that would allow for parallel processing and develop such a framework and the corresponding formalisms. Finally, though INGIT has shown some success in developing a domain specific implementation of a cross-modal translation system, its greatest success may be in raising some of the many representational and mapping problems that arise in such cross-modal translation.

However, as one of the first attempts to consider a semantic characterization for ISL and to have constructed a small prototype translation system, we hope that this work will lead to further exploration, both on the social and technological fronts, which would benefit the Indian deaf community.

Acknowledgement

We wish to thank AYJNIHH and Meher Dadabhoy for their help and Sujit and Sunil Sahasrabudhe for their inputs on Indian Sign Language.

References

- Alison Wray, Stephen Cox, Mike Lincoln and Judy Tryggvason, "A formulaic approach to translation at the post office: reading the signs", *Language and Communication*, 24: 59-75, 2004.
- A. L. Sexton, "Grammaticalization in American Sign Language", *Language Sciences*, 21: 105-141, 1999.
- Carl Pollard and Ivan E. Sag, *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago, 1994.
- Speers d'Armond, "Representation of American Sign Language for Machine Translation", Ph.D. Dissertation. Department of Linguistics, Georgetown University, 2002.
- Dilip Deshmukh, "Sign Language and Bilingualism in Deaf Education", India: Deaf Foundation, Ichalkaranji, 1996.
- Ian Marshall, Éva Sáfár, "A Prototype Text to British Sign Language (BSL) Translation System", *The Companion Volume to the Proceedings of 41st Annual Meeting of the Association for Computational Linguistics*, pages pp. 113-116, 2003.
- Joachim De Beule and Luc Steels, "Hierarchy in Fluid Construction Grammar", In: Furbach, Ulrich (ed.), *Proceedings of the 28th Annual German Conference on AI, KI 2005, Lecture Notes in Artificial Intelligence* (vol. 3698), pages 1-15, Berlin Heidelberg, 2005.
- Liwei Zhao, Karin Kipper, William Schuler, Christian Vogler, Norm Badler, Martha Palmer "A Machine Translation System from English to American Sign Language", *Proceedings of the Association for Machine Translation in the Americas 2000*, Published in *Lecture Notes in AI series of Springer-Verlag*, © Springer-Verlag, pages 54-67, 2000.
- Luc Steels and Joachim De Beule, "Unify and Merge in Fluid Construction Grammar", In: Lyon, C., Nehaniv, L. and A. Cangelosi (eds.), *Emergence and Evolution of Linguistic Communication, Lecture Notes in Computer Science*, Springer-Verlag: Berlin, 2006.
- Madan Vasishtha, James Woodward and Susan DeSantis, *An Introduction to Indian Sign Language*, All India Federation of the Deaf (Third Edition), 1998.
- Maria Papadogiorgaki, Nikos Grammalidis, Nikos Sarris and Michael G. Strintzis, "Synthesis of Virtual Reality Animations from SWML using MPEG-4 Body Animation Parameters", In *Proceedings Workshop on the Representation and Processing of Sign Languages - From SignWriting to Image Processing. 4th International Conference on Language Resources and Evaluation, LREC 2004*, pages pp. 43-50, 2004.
- Paul Kay and Charles J. Fillmore, "Grammatical constructions and linguistic generalizations: The whats x doing y? Construction.", *Language*, 75(1): 1-33, 2001.
- Rahul Banerjee and Amitabha Mukherjee, "Animating Hand Behaviours Using Virtual Sensors and an Automata Hierarchy", In *Proceedings Fourth Asian Conference on Industrial Automation and Robotics ACIAR-05*, May 11-13, 2005, Bangkok, Thailand, 2005
- Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hamke, and Jan Henning, *HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide*, volume 5 of *International Studies on Sign Language and Communication of the Deaf*. Signum Press, Hamburg, Germany, 1989.
- Tony Veale and Alan Conway, "Cross-Modal Comprehension in Zardoz, An English to Sign Language Translation system", presented at *The Fourth International Workshop on Natural Language Generation*, Maine, USA, 1994.
- Ulrike Zeshan, *Sign Language in Indopakistan: A Description of a Signed Language*, Amsterdam: John Benjamins, 2000.