# Converting Multilayer Glosses into Semantic and Pragmatic forms with GENLIS

**Rodolfo Delmonte**
Ca Foscari
University of Venice
Dept of Language Sciences
delmont@unive.it

**Serena Trolvi**
Ca Foscari
University of Venice
Dept. of Language Sciences
trolvi.serena@gmail.com

**Francesco Stiffoni**
Ca Foscari
University of Venice
Dept of Language Sciences
fstiffo@sgajo.com

**Abstract**

This paper presents work carried out to transform glosses of a fable in Italian Sign Language (LIS) into a text which is then read by a TTS synthesizer from an SSML modified version of the same text. Whereas many systems exist that generate sign language from a text, we decided to do the reverse operation and generate text from LIS. For that purpose we used a version of the fable *The Tortoise and the Hare*, signed and made available on Youtube by *ALBA cooperativa sociale*, which was annotated manually by second author for her master's thesis. In order to achieve our goal, we converted the multilayer glosses into linear Prolog terms to be fed to the generator. In the paper we focus on the main problems encountered in the transformation of the glosses into a semantically and pragmatically consistent representation. The main problems have been caused by the complexities of a text like a fable which requires coreference mechanisms and speech acts to be implemented in the representation which are often unexpressed and constitute implicit information.

## 1 Introduction

This paper presents work carried out for the automatic generation of written text in Italian language starting from glosses of fables in Italian Sign Language (LIS). The paper focuses on the semantic and pragmatic representation that has been created by the system GENLIS that feeds the generator. Whereas many systems exist that generate sign language from a text (Lombardo et al., 2011; Morrissey and Way, 2013; Wu et al., 2001; Sáfár and Marshall, 2001), we decided to do the reverse operation and generate text from LIS. A number of systems exist for American Sign Language that have attempted the same operation but only on a simple sentential basis and starting from visual recognition ((López-Ludeña et al., 2013; Dreuw et al., 2011; Elliott et al., 2000; Efthimiou et al., 2010)). The possibility to produce glosses automatically from video capture and image recognition (but see(Dorner and Hagen, 1994)) is not available for LIS, so we chose not to tackle the visual recognition phase and to start directly on the output, i.e. glosses[1]. Glosses for LIS are partly domain dependent in the sense that annotating sentences is a different task from annotating a dialogue, and this in turn is different from annotating a story or a fable. Among the many types of text that we could work on we chose the most difficult one: a fable, which is a mixture of narrative text and dialogues. For that purpose we used a version of the fable *The Tortoise and the Hare*, signed and kindly made available by *ALBA cooperativa sociale*. The signed story was annotated manually into glosses by second author - who is a LIS translator - for her Master's thesis (see also (Trolvi and Delmonte, 2020)). The fable has two main characters - and other secondary characters - with totally different personalities which may interact in dialogues, or may be simply narrated thus producing an overall complex textual structure.

## 2 Semantic and Pragmatic Representations from Glosses

As will be explained below, main problems have been caused by the complexities of a text like a fable - which is partly a dialogue and partly narration - and requires coreference mechanisms and speech acts to be implemented in order to convert glosses into a semantically and pragmatically consistent representations. The final text is organized into Discourse Units (hence DUs) or turns where each one may contain one or more sentences, and is associated with a unique turn identifier and a unique

[1]Transcription into glosses is a topic of research in itself because it may be done in different manners (Slobin et al., 2001; Hoiting and Slobin, 2002)

speaker. Eventually we came up with 30 DUs, 54 sentences and 91 propositions. The full project is presented on a website *https://genlis.vercel.app/*. The website contains full representations for the all the DUs of the fable, showing the conversion process step by step. Every DU starts by the video clip of the actor performing the LIS narration of the current DU; this is followed by the multilayered annotation[2] which is then turned into the 9 slot prolog consistent vector-like term. The transcribed vector is then enriched by semantic information and then by pragmatic information. The final step is the Italian sentences produced by the generator[3], which are then spoken aloud by the speech synthesizer on any Mac or PC. The final part of *GENLIS* addresses the speech synthesizer with a set of prosodic markers to induce correct pauses, voice volume, intonational movements[4]. For lack of space we cannot comment on this part of the system: we can only say that we are using SSML on available speech synthesizers that accept it, to produce an expressive and semantically correct recital of the story. State-of-the-art generation systems work mostly on the basis of a machine learning approach (Stein et al., 2012), (Zhao et al., 2000), which crucially requires an adequate amount of training data to feed the model. In our case training data are not available[5] also because glosses for LIS are partly domain dependent as said above. In our case, we decided to generate text from a LIS version of the fable **The Tortoise and the Hare** which has two main characters with totally different personalities. As will be clear from the sections below, we decided to follow a traditional approach which apart from the starting phase - content determination made available by the glosses - continues with text structuring, sentence aggregation, lexicalisation, referring expression generation, and linguistic realisation. These phases could also be understood as the sequence of processes of ATLAS project (Lesmo et al., 2011), which however had the opposite task – thus a reversed input-output, i.e. generating LIS from Italian texts.
Generating text from manual multilayer glosses is different from traditional NLG (Natural Language Generation). Generation from LIS glosses does not follow from well structured data-sets or knowledge basis, nor is there a plan in order to build logically well-formed representations (Gatt and Krahmer, 2017). Glosses are mainly sequences of lemmata with some indication of plural number, negation, quantifiers with agreed features, numbers, personal pronouns. But then verbal, nominal and adjectival expressions are just lemmata, auxiliaries are missing and the same applies to copulative verb ”to be” (Chesi et al., 2008). There are eight layers which specify type of speech act, presence of spatio-temporal location adverbs, role of current turn taker. They need to be collapsed and accounted for in the conversion phase in order to organize predicate-argument structures with all available information and converge towards a discourse level semantic and pragmatic representation.

GENLIS is written in the logic programming language Prolog (Gal et al., 1991; Mellish et al., 2006; Reiter, 2010), which makes available DCG (Definite Clause Grammar) rules together with Difference Lists to support text generation. The sequence of processes carried out by the system are represented in Figure 2 below.

Semantic forms are composed by main predicate, propositional attributes (such as e.g. mood, negation, verbal tense), arguments and adjuncts. Furthermore, each argument has its own internal structure. Semantic forms constitute the string that is eventually fed as input to the generator and then processed, in order to generate Italian sentences. We will describe below both the process of conversion of glosses into semantic forms and the structure of semantic forms. We will skip the first step in the whole process, which is producing the glosses and is done manually. As described in detail in another paper (Trolvi and Delmonte, 2020), manual glosses may contain arbitrarily many layers but they have the goal to interpret the signs in a shared manner. They are basically multi-layer text annotations written in tables, which can be done using one of the many software

[2]Manual annotation of simpler texts - either narrative or conversational - is not a highly time-consuming activity and can be carried out by an expert in a relatively short time.

[3]We are not aware of the existence of many generators for Italian (Lesmo et al., 2011) apart from the ones built by some of our collaborators (see (Delmonte and Bianchi, 1998; Delmonte and Pianta, 2008)) who were also partly the authors of a smaller version of the current one. The generator is now a general tool to generate most Italian sentence structures, and has been used in a number of other applications, like question-answering from a Discourse Model (see (Delmonte, 2000)).

[4]Intensive work on speech synthesis has been done in the past and also currently (see (Delmonte, 2016))

[5]Parallel corpora LIS-Italian text are available in a small number: besides ATLAS project (Lesmo et al., 2011), there is (Chesi et al., 2008) and (Barberis et al., 2011), none of which, however, will suit the genre requirements of the fable.

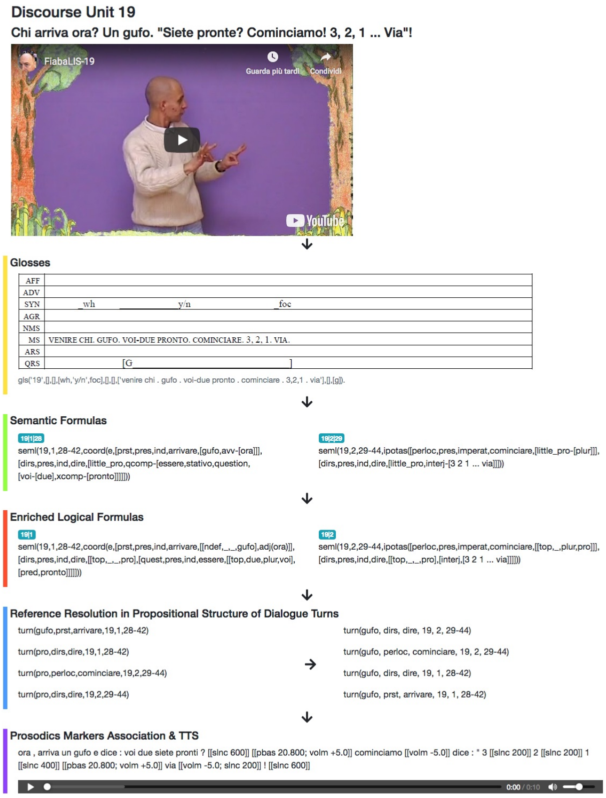


Figure 1: Snapshot of Discourse Unit 19 as presented by the website https://genlis.vercel.app dedicated to the generator.

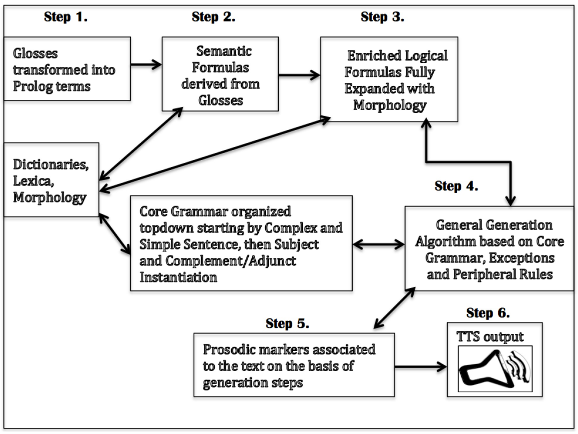


Figure 2: Flowchart of the GENLIS system decomposed into 6 steps.

available for the task - at first we used ELAN[6], but then we produced our own schemes to suit the requirements of the generator. In fact, in order for the multi-layer glosses to be analysed by the generator, it has been necessary to transform them into a 9 slot Prolog term. Thus, each annotation tier has been inserted in slots in a term, as follows:

[6]It can be downloaded here https://archive.mpi.nl/tla/elan - visited on April 2021

*gls(DUInd,Aff,Adv,Syn,Agr,Nms,Ms,Ars,Qrs)*

where the functor gls is an abbreviation for gloss, and contains a sequence of 9 slots explained here: DUind is the Discourse Unit index; the slots Aff, Adv, Syn contain annotated information about affective, adverbial and syntactic Non-Manual Signs; Agr identifies location and agreement of signs; Nms and Ms contain Non-Manual Signs and Manual Signs respectively and are expressed in a tokenized sequence between apostrophes ' ' as atomic objects; Ars and Qrs identify the occurrence of Action Role Shift and Quotation Role Shift[7].

## 2.1 The Conversion of Glosses into Syntactic/Semantic Lexical Forms

When creating conversion rules, we avoided indicating specific features that would make forms difficult to read and understand. More precisely, we did not indicate tense, mood and diathesis of verbs, number and genre of nouns and semantic role of oblique arguments for the generation of prepositions. We decided conventionally to generate sentences with active diathesis, in past tense and indicative mood. However, there are several factors to take into consideration: direct speech and questions, for example, are always expressed in present indicative. Furthermore, morphological features of nouns are always singular, unless otherwise indicated in glosses, and gender is derived from lexical gender. Past verb tense is derived on the basis of aspect of lexical verb; in particular, state and action verbs are expressed in *imperfetto* (a tense existing in Romance languages but not in English) tense, and the other verbs in past tense (*passato remoto* in Italian). Every fully expressed proposition has a verb that needs semantic and morphological features. While Person, Number and Gender may be inherited from the Subject, Tense and Mood are

[7]Role Shift is one of the main topics of the paper published on annotation of the fable (see (Trolvi and Delmonte, 2020)). It is a particular narrative strategy by which the signer adopts the perspective of another referent. Role Shift can be used to report a speech or thought of a referent or to reproduce his or her actions, thus it can be divided into two varieties. The terminology for both phenomena is not consistent throughout the literature. In our work, we adopted the terminology used by (Herrmann and Pendzich, 2018), namely "quotation role shift" (QRS) and "action role shift" (ARS). Hence, QRS is the type of RS by which the signer reports words or thoughts of other referents. ARS allows the iconic reproduction of actions, mannerisms and emotional states, including facial expressions and non linguistic gestures. It involved the use of the upper parts of the body (e.g. torso, head, eye gaze).

semantically and pragmatically determined. We have used lexical properties and discourse related (pragmatic) properties to assign Tense and Mood together with general consideration defined on the basis of narratological criteria. A fable or children story may be expressed using Indicative Present or Past tense (or *passato remoto*), however contextual conditions may impose constraints that require other Mood and Tense to be assigned. We may need to use Future tense, Imperative mood, Past tense (or *passato remoto*) rather than Present tense. A first subdivision of Mood-Tense assignment depending on Speech Act is shown below, a second subdivision follows according to Lexical Aspectual properties.

- Presentative constructions
  Perlocutive utterances
  Question + Exclamation
  Illocutive constructions
  Direct Speech constructions
  Statements

We distinguish Perlocutive from Illocutive verbs on the basis of the pragmatic nature of the action expressed: instructions on how to carry out a task are tagged Perlocutive and are enacted with Imperative mood. Illocutive expressions are tagged when the utterance expresses a decision or a wish to come true and are placed in the future Tense. Then, as a general rule, Activities are realized with Indicative Imperfetto, while Achievement use Past-tense (passato-remoto). The remaining cases are all realised with Indicative Present.

Semantic forms are structured as Prolog terms. Consistent with First-Order Logic (FOL), each term represents the content of a semantic proposition and is preceded by the functor PROP. PROP is the abbreviation for *proposition* and contains a fixed number of slots that mark semantic and pragmatic components included in glosses. More precisely, in first slot we may find pragmatic components like interjections - for expressing surprise or other affective and emotional aspects-, intrasentential elements like discourse markers and adverbs with scope on the verb or on the entire sentence. Let us now focus on the arguments structure. With the exception of SUBJect and OBJect, arguments are introduced by a functional marker that we derive from LFG theory (Bresnan, 2002), such as OBL for oblique arguments, FCOMP for sentential complement, VCOMP for verb complement and XCOMP for predicative complement. Moreover, argumental heads may contain modifiers, which are introduced by the marker MOD, or specifiers, which are usually included in brackets. If the argument is an expression of the affirmative or the negative polarity, the marker becomes the only term of the argument list. Moreover, direct speech is usually deprived of any introductory verb, which needs to be generated in Italian instead and may assume different meanings depending on context, as we will see in the next sections. Conversion rules from manual glosses are shown below:

- Identify elements that modify the main predicate, adverbs or discourse markers
- Insert the first verb you find
- Retrieve lexical verb aspect and create mood/time matrix
- The verb may be preceded by a location, which may be marked by a specific deictic term on the basis of type
- Speech act may vary
  PRESENTATION = WHO?
  DIRSPEECH for direct speech
  QUESTION if the sentence is a question
- Insert arguments into a list

**Generate nominal expressions:** The subject in first slot may be unexpressed. If so, it is marked with little-pro[8]: morphological features are retrieved from the subject of previous sentences. In case of direct speech, arguments may be interjections or statements/negations. The object may be a complement sentence marked FCOMP, an interrogative complement sentence marked QCOMP or an infinitive sentence marked VCOMP. Oblique arguments or adjuncts are marked OBL and in their first position they may contain either a preposition, if expressed overtly in manual glosses, or a semantic marker, and the lexical head in their second position. Nouns may have specifiers, such as *quale*/which in gara-[quale] (translated as race-[which]) and modifiers, which are marked MOD. Adverbs such as locative deictic adverbs are

[8]This label is derived from the Chomskyan linguistic theory that assumes the existence of an empty pronominal in pro-drop languages like Italian carrying morphological features derived from the main verb in sentences where the subject has been dropped, a choice which can be freely made in Italian and is based on discourse properties.

marked AVV. Gerundives are marked AVV too and contain the corresponding verb in infinitive form. PROPositions may be coordinated (COORD) or appear in sequence without markers (IPOTAS). These tags are inserted first, before the PROP tag. Examples are visible always in Figure 1 above.

All nominal expressions - both SUBJect and OBJect and OBLiques - can be modified by simple modifiers, multiple modifiers, and relative clauses. All of them are structurally attached to the nominal head because they are semantically and morphologically dependent on the head. In fact, adjectivals require feature agreement, which needs to be restricted before generation in order to prevent failures. As to relative clauses, their internal arguments may require the same type of information, in particular, in case the argument controlled by the relative pronoun - which may be unexpressed - is the SUBJect. Relative clauses may also be governed by an adjunct relation, but this is not the case in our story. In order to realise the appropriate word forms, the morphological features of the nominal head governing the relative clause are passed to the clause level as BINDER bundle of features, which may be used by the Verb Complex and realized as SUJBect or OBJect features.

**Generate Verbal Complex and Complementation:** The verb complex receives semantic and morphological information from the subject if present, be it a nominal or pronominal head, or simply an empty subject which however may have morphological features, person, number and possibly gender. Choosing the correct verbal complement structure may be dependent on subject semantic categories, which are also passed to the verbal complex. Semantic features are checked by matching subcategorisation information stored in the lexicon for each possible structural outcome. For instance, a verb like *dire*/say has a multiple entry in our computational lexicon with four different complement:

- vcomp = INFINITIVAL
  ogg = DIRECT-OBJECT
  ogg2 = INDIRECT-OBJECT(dative)+f/fcomp
  = SENTENTIAL-OBJECT
  f/fcomp = SENTENTIAL-OBJECT

They are all characterised by the same general lexical category, TRANSitive, and the same conceptual and semantic category, *report-dir* - that is a reporting verb that can be used also for direct speech introduction. This also applies to other verbs that may undergo intrasitivisation like mangiare/eat, but also to verbs with different complement structures but identical categorisations, like considerare/regard and dipingere/paint. In particular, *considerare*/consider has an open complements like NCOMP (a nominal predicative complement) or XCOMP (a label for generic open complements including infinitivals). All open complements require morphological features to match, and this will allow for complement structures to impose agreement for those features. This can be different for other verbs where lexical category may vary, as is the case for accennare/hint that may change from intransitive to transitive; or for a verb like apparire/appear that may change from copulative to unaccusative. Our lexicon is organised around a limited number of entries, around 1000 for most frequent lexical entries according to frequency dictionaries[9], and another extended set of manually annotated entries, around 9000, for the remaining less frequent but always non rare entries, which have a different feature and argument organization. Aspectual categories are very important - as said above - in the choice of verbal morphology regarding Tense and Mood; while semantic and conceptual class may also be relevant in case a sentential complement is present, as will be clarified below.
Another important feature of verbal complex is the requirements it poses on auxiliary choice and precise morphological information as to the Tense and Mood to be realised. In particular, simple vs. composite verbal complex may be realised, which in turn require specification of the appropriate auxiliary verb: *essere* for passive, reflexive, inherent reflexive and unaccusative classes, *avere* for active transitive and intransitive classes. Morphological information from the SUBJect is also required in case of auxiliary essere in order to generate the appropriate past participle. The same is required from the OBJect in case of pronominalization processes of the nominal head into a clitic pronoun, which however requires decisions that can only be made by a full-fledged pronoun resolution system - which is not implemented in the generator. As

[9]The list is derived from previous work on Italian Frequency Dictionaries, see (Delmonte et al., 1996)

to Person, this may be available in case the SUBJect is lexically expressed. Empty pronouns on the contrary do not realise Person feature, which is by default set to 3rd. Special cases are constituted by Imperative mood and Direct Speech. Imperative mood requires 2nd person to be realised if the command or instruction is addressed directly to the interlocutor. But there are commands in the fable addressed by the owl to both competitors, the hare and the tortoise, to start the race. In this second case, 2nd person plural is required. However, 1st person plural is also acceptable. Introducing 2nd person is not an easy task and we haven't been able yet to find a linguistically motivated trigger to do it. The verb is checked for agreement with SUBJect morphological features. This may cause failures in the generation step, until the appropriate verb form is produced.

Complements and adjuncts are selected according to the shape of the semantic form: nominal and sentential complements are made up of a four or five slots list, while an oblique may be constituted by a list containing five or six slots; a simple modifier has only two or three slots. Finally adverbials or interjections consist of one or two slots but contain a special label as unique identifiers. Sentential complements may be simple sentences preceded by a complementizer, which is locally generated; or they may be direct questions. In this second case, a question mark is added at the end. The two complement types are marked by a special label identifier FCOMP and QCOMP. A special case may be constituted by WH- questions as sentential complements, requiring a local WH- expression to be generated before the verb also in case it is an adjunct - i.e. when, how, where. These pronouns would be positioned after the verb in the logical form built from semantic forms. So they need to be raised, i.e. removed from the complement structure and generated in the appropriate position.

**Semantic Conversion Rules for Peripheral structural Representations:** Peripheral structures are those special stylistically marked structures, like Subject Locative Inversion with presentative structures, and complements realised as clitics, which need to be positioned before the verb. In both cases we implemented the rules to act at the end of the generation process. A SUBJect-Locative is used in the first sentence of the fable, when the hare is presented and appears on the scene as living in the woods. This is a typical introductory sentence for many fables or children stories and has all the required linguistic features: the protagonist is unknown and is realised as an indefinite nominal structure; the verb is unaccusative or intransitive. In this case *vivere*/live is used intransitively; the sentence is completed by presence of a Locative adjunct, *nel bosco/in the wood*. The main linguistic elements are all generated in their base structure, they are identified and displaced in order to produce a presentation structure where the Locative comes in first position followed by the verb complex and then comes the subject nominal and finally the rest of the sentence, which in this case is an apposition. The second rule of Inversion regards the well known Subject/Object Inversion in Direct Speech utterance where what is being said is positioned before the governing communication verb. For example, in the utterance DU.11 *Si', si', qui, rispose la lepre/Yes, yes, here, replied the hare* the generated sentence has the so-called deep order, Subj GovVerb Obj(the spoken utterance). The peripheral rule has the task to invert Obj and Subj and obtain the more naturally pronounced utterance, where the most important part (what is being said) comes at the beginning.

The second case of peripheral rule is the one involving the generation of a clitic pronoun *ci* for a locative or a dative repeated in the same complement structure, and the governing verb *partecipare*/participate. The clitic is generated after the verb and then it is scrambled before it. Structures that require special rules to be implemented include so-called Open Complements and Open Adjuncts. Open Complements are predicative complements of copulative verbs, as in *siete pronti*/are you ready; Open Adjuncts are state adjectives like *tranquillo*/quiet, which require gender/number agree-

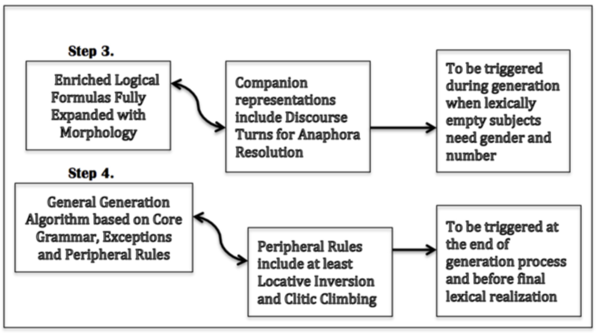


Figure 3: Peripheral Rules activated during Conversion and Generation.

ment with the SUBJect as in *la tartaruga guardava tranquilla*/the tortoise was watching quiet. Both cases require SUBJect morphological features to be visible in the Complement/Adjunct section of the generator in order to select or restrict the appropriate word form.

## 3 Special Rules required by Implicit Elements

There is a number of rules that need to be organized mainly inside the conversion portion of the system. These rules regard a number of specific features that are missing in the LIS glosses and in the sign language as a whole. They concern *Definiteness Assignment*, that is the need to add an article in Italian sentences in front of a nominal expression, which could also be zero article. Then there is the need to vary the direct speech introductory verb which is otherwise always reported as DIRE/say. Eventually there is the need to map Tense, Mood and Person/Gender/Number onto all verb complexes.

### 3.1 The algorithm for Definiteness Assignment

In order for the generation to work properly, the feature *definite*, *indefinite* or *zero* must be decided automatically and inserted in the list of features associated to each nominal expression, be it the primary head as with subjects and object, be it secondary with obliques where the noun phrase is governed by a preposition. The list of features includes morphological, semantic and informational features as follows:

*[Def,Spec,Num,Head]*

Def contains the information about definiteness if the head is a noun, otherwise it is substituted by TOP in case the head is a pronoun. Spec contains information on quantification and any linguistic element that may be expressed by a quantifier. Num is associated to the morphological feature of Number. The Algorithm for Definiteness Assignment (ADA) is based on two parameters: the type of constituent and the semantics associated with the noun. The semantics is taken from a set of different sources due to their dimensions, which are insufficient to cover all nominal expression of the fable. We have been using the lexical-semantic database ItalWordNet(see footnote below), and the list of semantic general categories annotated therein. In the algorithm the main call is known-def, which is used to memorize the type of definiteness associated to a nominal head. When a noun is met for the first time it is asserted as NDEF i.e. indefinite, unless it belongs to a set of exceptions and special semantic classes. The choice of zero definiteness applies to nominal expressions characterized by an abstract feature, which in ItalWordNet[10] is represented by MNT (= mental) and EXPR (= expressive) tags. It also applies to words indicating location tagged by PART (= part) and PLAC (= place). Another interesting class is constituted by words belonging to Body-Part like *orecchio/ear*, which are tagged as definite and characterized by features PART, LIV (= living) and FNCT (= function); the same applies to nouns belonging to TIME semantic class, like days, months, but also *appuntamento/date* whenever they are included in a nominal constituent. The list of these nominals in our fable includes the following words:

**appuntamento, vergogna, orecchio, giro, sinistra, destra, primo, tono/date, shame, ear, turn, left, right, first, tone**

In addition, glosses' expressions like *referente-N* where there is a number varying from 1 to 2, are treated as pronouns. Frozen expressions like *3 2 1 ... via/3,2,1...go* are marked with definiteness zero. The number belonging to the class of ordinals is tagged with zero definiteness only in case they are included in an oblique governed by *arrivare/come*. Of course, all adverbial like expressions and interjections are not considered and do not receive a list of morphological and semantic tags as said above.

### 3.2 The Algorithm for Narrative direct speech speaking verb type

Discourse level processing is the most complex part of the algorithm, because it is responsible for overall discourse coherence and cohesion. In the glosses, direct speech is introduced always by the same verb *dire/say*. It may also be deprived of any introductory verb, which in our case needs to take into account the semantic content of the current utterance. In addition, depending on current discourse turn speaker, this verb may assume different meanings, which are strictly discourse

[10]https://www.cnr.it/it/banche-dati-istituti/banca-dati/442/italwordnet-iwn

related. So either *dire/say* is substituted by a contextually determined verb or a verb is introduced which was not present. These verbs belong to the Answering semantic type and are: *rispondere/reply* or *replicare/reply*, in case the speaker is answering a question from previous discourse turn. Otherwise the predicate may belong to the Asking type, *chiedere/ask* or *domandare/ask*, in case the current turn is made of a question; eventually it may also be *dire/say* in case the previous turn was a yes/no question, or the current turn is a statement. Finally with exclamations it may be *esclamare/exclamate*. The algorithm is part of the convert file, the conversion algorithm that starting from glosses organizes them into semantic forms. It is activated after all conversions have been already made. The call is intended to modify the current predicate in case it is needed by the context. This is done checking semantic forms. Each turn is a vector representation, with current topic speaker, current speech act associated to current utterance, and a main predicate. The main predicate is headed by a Discourse Unit index, a Sentence index and a proposition index, like this: Head-Spac-Pred-Du-Sn-N. These representations are asserted into memory in a Prolog database and may be extracted easily.
The conversion algorithm receives Semantic Forms and checks to verify whether the current verb is *dire/say*. It also contains the current governing predicate, the arguments of current predicate in the Body variable, the Arguments of the first sentential complement (if any) of the Body variable in the variable Args, and finally the variable NewBody that will contain the modified version of the arguments. The first call to modify the predicate checks to see what is the speech act of the first proposition chosen. In this case the Predicate is substituted by a predicate of the Asking type, *chiedere-domandare/ask*. The second call is the most important one and is accompanied by a check of the previous turn. The call to verify previous turns is used to look into the database of turns. The search is interrupted in case the current utterance contains a question as one of its sentential complements. Then the second call searches the turns database. At first it extracts the previous turn and then it checks to see whether the current topic is different from the one asserted in the previous turn; finally it checks whether the speech act is a question. In this case the main predicate is modified into one of the Asking type. Eventually, the output of the generator is semantically coherent and pragmatically correct but it is fairly different from the one we created to stylistically suit a typical fable and interpreting the signer. Consider for instance the output of the generator for *DU 19: Ora arriva un gufo e dice : voi due siete pronti ? 3 2 1 via./Now comes an owl and says: you two are ready ? 3 2 1 go*. Compared to the utterance manually built corresponding to stylistically suit a typical fable story: *Chi viene ora? Un gufo. Siete pronte? Cominciamo! 3 2 1 ... via!/Who is coming now? An owl. Are you ready/fem/plur? Let's start! 3 2 1 ... Go!* This is shown in the figure below which is an excerpt from the website:

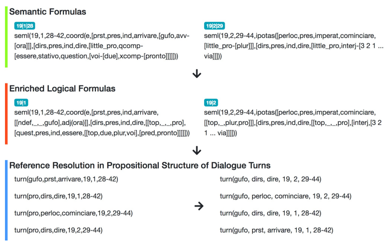


Figure 4: Excerpt of DU n.19 showing only the semantic and pragmatic conversion steps.

## 4 Evaluation and Discussion

Evaluation can be done manually or automatically (Belz and Reiter, 2006; Novikova et al., 2017). In order to do it automatically one would need a corpus of fables to be used for training which we currently don't have available. One should also take into account the need to measure how well glosses for LIS have been created and have been used by the system to produce a naturally sounding Italian text which resembles a fable. Also this evaluation is difficult to make for the same reason. We turned to human evaluation for lack of a better opportunity now left for the future. In order to evaluate the output of the generator we wrote manually a version of the story which was more adherent to what is expected from children fables and is attested in online versions of this fable. At the same time, the made up version had to respect as faithfully as possible the signed version produced by the signer in the video. The result is a story which is pleasant to listen to by children and adults, as we tested in a primary school classroom for an experiment. Now comes the evaluation of the generated story that we are able to produce by a comparison with the manually created story - that we make available in full in the supplementary materials. The comparison was done at the beginning in order to produce the peripheral rules presented in the section above. What has been left unchanged is discussed here below. We decided to grade each Discourse Unit or SubUnit by a four levels graded scale: 1 = No Difference, 2 = Slight Differences, 3 = Noticeable differences, 4 = Very different.

**1 = No Difference**

No Discourse Unit or SubUnit is totally identical

**2 = Slight Differences**

a) Definiteness Assignment in DU. 1 un $\neq$ il, una $\neq$ la

P. In un bosco viveva una lepre, una lepre altezzosa.

G. Nel bosco viveva una lepre la lepre altezzosa.

We discuss this point using the first Discourse Unit of the story which we show here in the P(roposed) form and the G(enerated) form. The rule we created regards certain words as generic nominals which do not need to be individuated in the world and are assigned definiteness as they appear. This is the case of *bosco/wood*. The case of *una lepre/a hare* is different: at first appearance the nominal *lepre* is correctly assigned an indefinite article (una/a); as to second appearance our system computes HARE as already known in the world and assigns definiteness (la/the). But in this case the syntactic function of apposition reverts the semantics, because the apposition is just a means of characterizing the entity with additional attributes or properties. However this is difficult to realize in the generator.

DU1, DU7

b) Different Mood/Tense Present vs. Past Tense in DU. 2 avvicinò / avvicina

P. La lepre le si avvicinò ...

G. La lepre si avvicina ...

The rule for Mood/Tense assignment is sensitive to aspectual classes and speech act and we don't have the possibility to revert Present Tense to Past Tense in this case

DU.2, DU.3, DU4, DU5, DU8, DU15, DU18.2, DU22, DU23.1, DU23.2, DU25, DU26, DU27.2, DU28, DU29, DU30

c) Dative Ethic in DU. 2 le??

P. La lepre le si avvicinò ...

G. La lepre si avvicina ...

Presence of a Dative Ethic in Italian is optional and does not contribute to modify the semantics. We don't know of a linguistically motivated rule which could be used to insert it and make the sentence sound more natural

DU2

d) Use of a different direct speech communicative verb domandò/chiede

P. La tartaruga perplessa domandò...

R. La tartaruga chiede con aria perplessa...

DU3, DU4, DU5, DU9, DU10, DU11, DU12, DU17, DU25

e) Use of a different wh- word Che $\neq$ quale

DU6,

f) Use of a different locative adverbial lì $\neq$ qua in fondo $\neq$ là

DU7.1, DU7.2, DU27.1

g) Use of a different but fully synonymous verb from the one signed and inserted in the glosses

DU7.3, DU16, DU23.2, DU27.1

h) Use of a different exclamation interjection from the one signed

DU12

i) Presence of additional material in the generated story which was however present in the glosses and has been erased by the manually created story because redundant

DU13, DU14, DU27.3, DU29, DU30

l) Deletion of governing communicative verb in the generated story

DU10.2

**3 = Noticeable differences**

a) Omission of predicates present in the glosses

DU16.2

b) Omission of linguistic material like personal pronouns needed to reinforce the assertion

DU16.2, DU18.2

c) Mistaken gender associated to subject noun phrase or predicative open complement in copulative structures I-masc-plural due/Le-fem-plural

due
P. Le due si affiancarono.
G. I due si affiancarono.
DU18.1

d) Insertion of additional linguistic material in the manual story which was not present in the glosses DU1.2, DU7.1, DU18.2, DU18.3, DU19, DU20, DU21, DU23.1, DU27.2, DU29

e) Presence of linguistic material which is semantically almost synonymous but lexically different from the one proposed in the glosses DU20, DU23

f) Presence of identical Noun Phrase in two coordinated sentences which sounds redundant and should have been pronominalized as has been done in the manual story
DU22, DU24

Eventually, we recorded no case of identical utterances, 8 cases of Noticeable Differences due to our algorithm and a higher number (10 cases) of arbitrary or stylistically motivated insertion of linguistic material in the manual story. The remaining mismatches (45) are to be regarded minor or Slight Differences which should be corrected in the future by further developments of the main algorithm. Overall, on a total of 54 Sentences and 91 simple sentences or propositions, we had 63 mismatches only 8 of which had a semantic impact on the story, which amounts to less than 10% error rate.

## 5 Conclusion

In this paper we presented the conversion process produced by *GENLIS*, a system that generates Italian text from glosses of the Italian Sign Language (LIS). The signed text we chose is a fable, i.e. a semantically and pragmatically difficult text to generate. We described all the steps that are required to convert a vector-like representation of the multi-layered annotation scheme used for transcribing signs into glosses. To complete our experiment, we did an evaluation by comparing the output of the generator to a manually written version of the story to suit stylistic requirements for fables and came to the conclusion that the result is acceptable but for a few particularly difficult utterances. Eventually we only had 8 semantically relevant mismatches over 63 as a whole. However we had to overcome a number of problematic issues at a morphological, syntactic and semantic level which were successfully solved thanks to peripheral rules executed at the end of the generation process. Future work includes improving the algorithm to generate a story which is more natural and pleasant to listen to. It shall also address separately either dialogues or narrative texts in order to produce a consistent and more generalized conversion process from glosses to spoken utterances.

## References


D. Barberis, N. Garazzino, P. Prinetto, G. Tiotto, A. Savino, U. Shoaib, and N. Ahmad. 2011. Language resources for computer assisted translation from italian to italian sign language of deaf people. In *Proceedings of Accessibility Reaching Everywhere AEGIS Workshop and International Conference*, Brussel.

Anja Belz and Ehud Reiter. 2006. Comparing automatic and human evaluation of nlg systems. in proceedings of the 11th conference of the european chapter of the association for computational linguistics. pages 313–320, Trento.

Joan Bresnan. 2002. *Lexical-Functional Syntax*. Blackwells.

Cristiano Chesi, Gianluca Lebani, and Margherita Pallottino. 2008. A bilingual treebank (ita-lis) suitable for machine translation: what cartography and minimalism teach us. *STUDIES IN LINGUISTICS*, 2:165–185.

Rodolfo Delmonte. 2000. Generating from a discourse model. In *Proceedings of the MT 2000 - MACHINE TRANSLATION AND MULTILINGUAL APPLICATIONS IN THE NEW MILLENNIUM*, pages 313–320, Exeter(UK). British Computer Society - BCS.

Rodolfo Delmonte. 2016. Expressivity in tts from semantics and pragmatics. In *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*, pages 407–427, Milano, Italy.

Rodolfo Delmonte and Dario Bianchi. 1998. Dialogues from texts: How to generate answers from a discourse model. In *Atti Convegno Nazionale AI*IA*, page 139–143.

Rodolfo Delmonte, Giacomo Ferrari, Anna Goy, Leonardo Lesmo, Bernardo Magnini, Emanuele Pianta, Oliviero Stock, and Carlo Strapparava. 1996. Ilex - un dizionario computazionale dell'italiano. In *Proceedings of 5th Convegno Nazionale della Associazione Italiana per l'Intelligenza Artificiale - AI*IA "Cibernetica e Machine Learning"*, pages 27–30, Napoli, Italy.

Rodolfo Delmonte and Emananuele Pianta. 2008. Answering why-questions in closed domains from a discourse model. In *Proceedings of Semantics in Text Processing (STEP)*, page 109–114.

Brigitte Dorner and E. Hagen. 1994. Towards an american sign language interface. *Artificial Intelligence Review*, 8:235–253.

P. Dreuw, J. Forster, Y. Gweth, D. Stein, H. Ney, G. Martinez, J. V. Llahi, O. Crasborn, E. Ormel, W. Du, T. Hoyoux, J. Piater, J. M. Moya, and M. Wheatley. 2011. Signspeak – understanding, recognition, and translation of sign languages. In *Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 22–23.

E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, and F. Goudenove. 2010. Dicta-sign: Sign language recognition, generation and modelling with application in deaf communication. In *CSLT 2010 - LREC 2010*, pages 80–83.

R. Elliott, J.R.W. Glauert, J.R. Kennaway, and I. Marshall. 2000. The development of language processing support for the visicast project. In *4th International ACM SIGCAPH Conference on Assistive Technologies (ASSETS 2000)*, pages 101–108.

A. Gal, G. Lapalme, P. Saint-Dizier, and H. Somers. 1991. *Prolog for Natural Language Processing*. John Wiley Sons Ldt, Chinchester.

A. Gatt and E. Krahmer. 2017. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170.

A. Herrmann and N.-K. Pendzich. 2018. Between narrator and protagonist in fables of german sign language. *Linguistic foundation of narration in spoken and sign languages*, pages 275–308.

N. Hoiting and D.I. Slobin. 2002. *Transcription As A Tool For Understanding: The Berkeley Transcription System For Sign Language Research (BTS)*, pages 55–75. John Benjamins, Amsterdam/Philadelphia.

Leonardo Lesmo, Alessandro Mazzei, and Daniele Radicioni. 2011. Linguistic processing in the atlas project. In *Proceeding of International Workshop on Sign Language Translation and Avatar Technology(SLALT)*.

Vincenzo Lombardo, Cristina Battaglino, Rossana Damiano, and Fabrizio Nunnari. 2011. An avatar-based interface for the italian sign language. In *International Conference on Complex, Intelligent and Software Intensive Systems, CISIS 2011, June 30 - July 2, 2011, Korean Bible University, Seoul, Korea*, pages 589–594. IEEE Computer Society.

Verónica López-Ludeña, Roberto Barra-Chicote, Syaheerah Lutfi, Juan Manuel Montero, and Rubén San-Segundo. 2013. Lsespeak: A spoken language generator for deaf people. *Expert Systems with Applications*, 40:1283–1295.

C. Mellish, D. Scott, L. Cahill, D. S. Paiva, R. Evans, and M. Reape. 2006. A reference architecture for natural language generation systems. *Natural Language Engineering*, 12(01):1–34.

S. Morrissey and A. Way. 2013. Manual labour: tackling machine translation for sign languages. *Machine Translation*, 27(1):25–64.

Jekaterina Novikova, Ondrej Dusek, Amanda Cercas Curry, and Verena Rieser. 2017. Why we need new evaluation metrics for nlg. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

E. Reiter. 2010. *Natural Language Generation*, pages 574–598. Wiley-Blackwell.

D. I. Slobin, N. Hoiting, M. Anthony, Y. Biederman, M. Kuntze, R. Lindert, J. Pyers, H. Thumann, and A. Weinberg. 2001. Sign language transcription at the level of meaning components: The berkeley transcription system (bts). *Sign Language Linguistics*, 4:63–96.

D. Stein, C. Schmidt, and H. Ney. 2012. Analysis, preparation, and optimization of statistical sign language machine translation. *Machine Translation*, 26(4):325–357.

E. Sáfár and I. Marshall. 2001. The architecture of an english-text-to-sign-languages translation system. In *Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing (RANLP-02)*, pages 223–228.

Serena Trolvi and Rodolfo Delmonte. 2020. Annotating a fable in italian sign language (lis). In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 6025–6034, Marseille, France. European Language Resources Association.

C. H. Wu, H. Y. Su, Y. H. Chiu, and C. H. Lin. 2001. Transfer-based statistical translation of taiwanese sign language using pcfg. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6.

L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, and M. Palmer. 2000. A machine translation system from english to american sign language. In *Envisioning Machine Translation in the Information Future: Proceedings of the Fourth Conference of the Association for Machine Translation (AMTA-00)*, pages 293–300.