

Manual for XLE+Glue system

Mary Dalrymple¹ and Agnieszka Patejuk^{1,2} and Mark-Matthias Zymla³

¹Centre for Linguistics and Philology, University of Oxford

²Institute of Computer Science, Polish Academy of Sciences

³University of Konstanz

mary.dalrymple@ling-phil.ox.ac.uk

agnieszka.patejuk@gmail.com

mark-matthias.zymla@uni-konstanz.de

Contents

1	Overview	3
2	Encoding of glue premises as AVMs with embedding	4
2.1	Meaning constructors and their AVM encoding	4
2.2	Universal quantification over meaning constructors	6
2.3	Multiple meaning constructors contributed by a single word . . .	7
2.4	Templates for meaning constructors	8
2.4.1	The basic definitions	8
2.4.2	Non-implicational meaning constructors: Proper names .	8
2.4.3	Meaning constructors requiring arguments: Intransitive verbs	9
2.4.4	Scope of local names	10
3	Sample grammars	11
4	Formatting for glue premises: General representational issues	11
4.1	Using CONCAT	11
4.2	Special characters in XLE	12
4.3	Using backquote-region.el	12
5	Semantic representations and the prover	12
5.1	Parsing linear logic formulas	12
5.2	System without the semantic parser	13
5.3	Using the semantic parser	13
5.3.1	Lambda expressions	13
5.3.2	Complex types	14
5.3.3	Logic operators	14
5.3.4	Quantifiers	14
5.3.5	Functional application	14
5.3.6	Lexical entries in XLE	15

6	Prolog rewrite component	15
7	Illustrating the flexibility of the system	16
7.1	Meaning constructors encoded as strings	16
8	Setting up and running the system	17
A	Appendix: Running XLE in Windows 10 Subsystem for Linux	18

The XLE+Glue pipeline integrates functionalities from various programming languages. First, Tcl provides the functionality for the XLE user interface. We provide an `xlerc` and a Tcl file that set up all components of the pipeline: the grammar, the rewrite component, and the GSWB. The UI is modified to allow the user to call the XLE+Glue system from the XLE user interface. Second, SWI-Prolog is used to rewrite the AVM encoding of meaning constructors into the appropriate string format.¹ Third, Java in which the GSWB is implemented.

<https://github.com/Mmaz1988/xle-glueworkbench-interface>

1 Overview

The diagram illustrates the GlueSemWorkbench architecture. It starts with an **XLE+Glue Grammar** and a **Parser**, which feed into a central processing unit. This unit has three input paths: **Prolog rewrite**, **Prolog f-structure**, and **XLE User interface**. The central unit is labeled **GlueSemWorkbench** and contains an **INPUT** section. The **INPUT** section shows a partial f-structure for the sentence "Kim smiled":

```

{
  Kim: 1_e
  smile: { 1_e -o 3_t }
}

```

The **GlueSemWorkbench** produces an **OUTPUT** section, which includes:

- Prolog:** 3 : [smile][Kim] [0 , 1]
- Sequent:**
 - 1 : Kim [0]
 - (1 -o 3) : smile [1]
- Agenda:**
 - 1 : Kim [0]
 - (1 -o 3) : smile [1]

Below the main diagram, a sample XLE f-structure for the sentence "Kim smiled" is shown:

```

[ PRED 'smile:[1Kim]>'
  HUMJ [
    GLUE [
      [ PRED 'Kim'
        ARG1 {
          [ ARG1 [SMOGR [1Kim]]
            TYPE o
          ]
          [ ARG2 [SMOGR [3smile]]
            TYPE t
          ]
        ]
      ]
    ]
  ]
]ENSE OUT, TYPE FIN

```

¹We did not use the XLE transfer system, since it is not available for newer XLE versions.

Glue meaning constructors consist of a semantic side (any semantic formalism) and glue side (expression of linear logic over linguistic structures of a given type). F-structure metavariables in lexical entries are instantiated to indices representing particular f-structures, for instance:

- (1) a. $Kim : \uparrow_e \Rightarrow Kim : k_e$
- b. $smile : (\uparrow \text{SUBJ})_e \multimap \uparrow_t \Rightarrow smile : k_e \multimap s_t$

In the first part of this manual we make several assumptions for ease of presentation. First, we assume simple untyped meaning representations such as *Kim* and *smile*; see Section 4 for discussion of the ways in which meanings can be represented in the system, including as terms of the typed lambda calculus. Second, we assume that the glue side of meaning constructors refers to f-structures rather than their semantic projections (as in our sample grammar `glue-basic.lfg`, described in Section 3). However, it is also possible for meaning constructors to refer to other linguistic structures in our system (as in our sample grammar `glue-basic-semstr.lfg`, which includes a semantic projection).

Our system provides two ways of encoding glue premises. Both methods rely on the presence of a special GLUE attribute whose value is a set of meaning constructors in AVM format. The pipeline shown on the left-hand side of the above diagram illustrates the AVM encoding: glue premises encoded in an attribute-value matrix (AVM) format are converted to the GSWB prover input format. We describe this method in Section 2. In the string-based method, the meaning constructor is represented as a sequence of substrings which are values of an ordered set of attributes, as we describe in Section 7.1. In both encodings, each meaning constructor is rewritten from the attribute-value format to a format suitable for input to GSWB. The Prolog rewrite script and GSWB run as separate processes, integrated into XLE. As a result, XLE+Glue allows the user to call the semantic analysis directly from the XLE output windows.

2 Encoding of glue premises as AVMs with embedding

2.1 Meaning constructors and their AVM encoding

(2-a) illustrates the meaning constructor for the proper name *Kim* in the standard format, and (2-b) illustrates the corresponding attribute-value encoding, where the AVM encoding the meaning constructor appears as a member of the GLUE set. In (2) and the rest of this section, we follow the usual notational convention of referring to f-structures by means of letters like *k* and *s*, but in later sections of this document we will use numbers instead, since the Prolog rewrite script relies on the numeric indices assigned to f-structures in the Prolog output format of XLE. In the examples in this section, each GLUE set contains only one meaning constructor, but in other cases several meaning constructors appear as members of the GLUE set, as we show in Section 2.3.

- (2) a. $Kim: k_e$

$$\text{b. } k : \left[\begin{array}{ll} \text{PRED} & \text{'KIM'} \\ \text{GLUE} & \left\{ \left[\begin{array}{ll} \text{MR} & \text{KIM} \\ \text{SEMSTR} & k \\ \text{TYPE} & e \end{array} \right] \right\} \end{array} \right]$$

In the attribute-value encoding, each glue meaning constructor minimally consists of three attributes. The value of MR is the semantic (left-hand) side of the meaning constructor, while SEMSTR and TYPE specify the glue (right-hand) side: SEMSTR points to the relevant f-structure, while TYPE specifies SEMSTR's type (*e* or *t*). In XLE notation, the f-structure constraints contributed by a proper name like “Kim” are:

$$\begin{aligned} (3) \quad & (\text{~ PRED}) = \text{'Kim'} \\ & (\text{\%sstr MR}) = \text{Kim} \\ & (\text{\%sstr SEMSTR}) = \text{~} \\ & (\text{\%sstr TYPE}) = e \\ & \text{\%sstr \$} (\text{~ GLUE}) \end{aligned}$$

`%sstr` is a local name² used to construct an f-structure containing attributes specifying the glue meaning constructor (MR, SEMSTR, TYPE). This f-structure is added to the GLUE set by specification of the constraint `%sstr $ (~ GLUE)`.

A glue meaning constructor may also involve implication, as for a verb like *smile* in (4), where a resource is consumed in order to produce another resource. The standard meaning constructor for the verb *smile* is given in (4-a), and the AVM translation is given in (4-b). In the AVM encoding, ARG1 is the first resource to be consumed, ARG2 the second, etc. The resources to be consumed are specified using the SEMSTR and TYPE attributes.

$$\begin{aligned} (4) \quad \text{a. } & \textit{smile}: k_e \multimap s_t \\ \text{b. } & s : \left[\begin{array}{ll} \text{PRED} & \text{'SMILE<SUBJ>'} \\ \text{SUBJ} & k : [] \\ \text{GLUE} & \left\{ \left[\begin{array}{ll} \text{MR} & \text{SMILE} \\ \text{ARG1} & \left[\begin{array}{ll} \text{SEMSTR} & k \\ \text{TYPE} & e \end{array} \right] \\ \text{SEMSTR} & s \\ \text{TYPE} & t \end{array} \right] \right\} \end{array} \right] \end{aligned}$$

This means that the constraints contributed by the verb *smile* are:

$$\begin{aligned} (5) \quad & (\text{~ PRED}) = \text{'smile<(~ SUBJ)>'} \\ & (\text{\%sstr MR}) = \text{smile} \\ & (\text{\%sstr SEMSTR}) = \text{~} \\ & (\text{\%sstr TYPE}) = t \\ & (\text{\%sstr ARG1 SEMSTR}) = (\text{~ SUBJ}) \\ & (\text{\%sstr ARG1 TYPE}) = e \\ & \text{\%sstr \$} (\text{~ GLUE}) \end{aligned}$$

The f-structure for the sentence “Kim smiles” is shown in (6). The Prolog rewrite component collects up all of the premises in the GLUE sets, rewrites

²<https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/notations.html#N4.1.6>. See the discussion of scope of local names in 2.4.4.

each premise into a format suitable for input to the prover (as shown in (7)), and passes the complete set of premises to the prover.

$$(6) \quad s : \left[\begin{array}{l} \text{PRED} \quad \text{'SMILE<SUBJ>'} \\ \text{SUBJ} \quad k : \left[\begin{array}{l} \text{PRED} \quad \text{'KIM'} \\ \text{GLUE} \quad \left\{ \left[\begin{array}{l} \text{MR} \quad \text{KIM} \\ \text{SEMSTR} \quad k \\ \text{TYPE} \quad e \end{array} \right] \right\} \end{array} \right] \\ \text{GLUE} \quad \left\{ \left[\begin{array}{l} \text{MR} \quad \text{SMILE} \\ \text{ARG1} \quad \left[\begin{array}{l} \text{SEMSTR} \quad k \\ \text{TYPE} \quad e \end{array} \right] \\ \text{SEMSTR} \quad s \\ \text{TYPE} \quad t \end{array} \right] \right\} \end{array} \right]$$

$$(7) \quad \left\{ \begin{array}{l} \text{Kim} : \text{k_e} \\ \text{smile} : \text{k_e} \multimap \text{s_t} \end{array} \right\}$$

2.2 Universal quantification over meaning constructors

Example (8) illustrates the use of the attribute `FORALL`³ for universal quantification in meaning constructors.

$$(8) \quad \text{a.} \quad \text{every: } \forall F. (p_e \multimap p_t) \multimap (m_e \multimap F_t) \multimap F_t$$

$$\text{b.} \quad m : \left[\begin{array}{l} \text{PRED} \quad \boxed{1} \, p \\ \text{GLUE} \quad \left\{ e : \left[\begin{array}{l} \text{MR} \quad \text{every} \\ \text{ARG1} \quad \left[\begin{array}{l} \text{SEMSTR} \quad \boxed{1} \\ \text{TYPE} \quad e \end{array} \right] \\ \text{SEMSTR} \quad \boxed{1} \\ \text{TYPE} \quad t \end{array} \right] \\ \text{ARG2} \quad \left[\begin{array}{l} \text{SEMSTR} \quad m \\ \text{TYPE} \quad e \end{array} \right] \\ \text{FORALL} \quad F \\ \text{SEMSTR} \quad F \\ \text{TYPE} \quad t \end{array} \right] \right\} \end{array} \right]$$

In (8)b, the f-structure labeled e is the attribute-value encoding of the meaning constructor for “every” given in (8)a. It has two arguments: `ARG1` represents the restriction of the quantifier, and `ARG2` represents its scope. The value of the `ARG1` attribute encodes the implication $(p_e \multimap p_t)$ (where p is the value of the `PRED` attribute of the noun phrase, as shown in (8)), which corresponds to a common noun meaning.⁴ The value of the `ARG2` attribute encodes an

³To display the attribute `FORALL` in XLE, select “constraints” in “Views” menu (or press “c”) in the window containing the glue premises in AVM format.

⁴ Our sample grammar `glue-basic.lfg` makes the non-standard assumption that the

implication from m_e to F_t , where F is a variable bound by a universal quantifier, representing the scope of the quantifier, which is freely chosen. At the top level we have a new attribute **FORALL**, which encodes the universal quantifier \forall in (8)a.

2.3 Multiple meaning constructors contributed by a single word

Example (9) illustrates the encoding of the meaning constructor for the quantifier *everyone*, decomposed into a meaning constructor contributing the “every” part of the meaning and another meaning constructor contributing the “person” part. This allows for the modification of the restriction of the quantifier in examples like *everyone who smiled*, and illustrates the possibility for a single word to contribute more than one meaning constructor to the GLUE set.

$$(9) \quad \text{a.} \quad \text{every: } \forall F.(p_e \multimap p_t) \multimap (m_e \multimap F_t) \multimap F_t$$

$$\text{person: } p_e \multimap p_t$$

$$\text{b. } m : \left[\begin{array}{c} \text{PRED} \quad \boxed{1} \text{ everyone} \\ \text{GLUE} \left\{ \begin{array}{l} \left[\begin{array}{l} \text{MR} \quad \text{every} \\ \text{ARG1} \left[\begin{array}{l} \text{ARG1} \left[\begin{array}{l} \text{SEMSTR} \quad \boxed{1} \\ \text{TYPE} \quad e \end{array} \right] \\ \text{SEMSTR} \quad \boxed{1} \\ \text{TYPE} \quad t \end{array} \right] \\ \text{ARG2} \left[\begin{array}{l} \text{ARG1} \left[\begin{array}{l} \text{SEMSTR} \quad m \\ \text{TYPE} \quad e \end{array} \right] \\ \text{SEMSTR} \quad F \\ \text{TYPE} \quad t \end{array} \right] \\ \text{FORALL} \quad F \\ \text{SEMSTR} \quad F \\ \text{TYPE} \quad t \end{array} \right] \\ \left[\begin{array}{l} \text{MR} \quad \text{person} \\ \text{ARG1} \left[\begin{array}{l} \text{SEMSTR} \quad \boxed{1} \\ \text{TYPE} \quad e \end{array} \right] \\ \text{SEMSTR} \quad \boxed{1} \\ \text{TYPE} \quad t \end{array} \right] \end{array} \right\} \end{array} \right]$$

In (9)b, the f-structure labeled e is the same as the attribute-value encoding of the meaning constructor for *every* given in (8) in the previous section. The second member of the set, labeled n , is the same as the contribution we would expect for the common noun *person*.

meaning of a common noun is a function from its **PRED** value of type e to its **PRED** value of type t ; that is, a common noun like “person” has a lexical entry of the following form:

$$\text{person: } (\uparrow \text{PRED})_e \multimap (\uparrow \text{PRED})_t$$

This is done for simplicity, to avoid the introduction of attributes encoding **VAR** and **RESTR** as in standard treatments, and is not a necessary feature of the implementation. We discuss our sample grammars in more detail in Section 3.

2.4 Templates for meaning constructors

The sample grammar files `glue-basic.lfg`, `glue-basic-semstr.lfg`, and `glue-basic-semparser.lfg` provide a set of templates⁵ for encoding meaning constructors which may be generally useful, though it is of course possible for grammar writers to develop their own set of templates or to modify the sample templates as needed. We describe the basic templates here; more discussion of the sample grammars can be found in Section 3 and in the comments in the grammar files.

2.4.1 The basic definitions

All of the templates which are used in defining meaning constructors in the sample grammars call the two basic templates `GLUE-SEMSTR` and `GLUE-MEANING`. The template `GLUE-SEMSTR` specifies an attribute-value structure `SSTR` as having the value `FSTR` for the attribute `SEMSTR` and the value `TY` for the attribute `TYPE`. The attributes `SEMSTR` and `TYPE` and their values must appear in all AVM meaning constructors and argument specifications, to identify the relevant linguistic resource and its type. (The argument names `FSTR` and `SSTR` are merely mnemonic; the linguistic resource need not be an f-structure.)

$$(10) \quad \text{GLUE-SEMSTR}(\text{FSTR SSTR TY}) = (\text{SSTR SEMSTR}) = \text{FSTR} \\ (\text{SSTR TYPE}) = \text{TY}.$$

The template `GLUE-MEANING` specifies an attribute-value structure `SSTR` as having the value `P` for the attribute `MR`. This attribute corresponds to the left-hand (meaning) side of the meaning constructor, and must appear once, at the top level of all AVM meaning constructors.

$$(11) \quad \text{GLUE-MEANING}(\text{SSTR P}) = (\text{SSTR MR}) = \text{P}.$$

2.4.2 Non-implicational meaning constructors: Proper names

In the sample grammar `glue-basic.lfg`, the lexical entry for the proper name *Kim* is as in (12):

$$(12) \quad \text{Kim} \quad \text{N} \quad * \quad @(\text{PROPERNOUN Kim}).$$

The sample grammar `glue-basic.lfg` provides the template in (13) for proper names. It defines the f-structure `PRED` value, and calls the template `GLUE-PROPERNOUN` to define the meaning constructor in AVM format, passing in the argument `P`.

$$(13) \quad \text{PROPERNOUN}(\text{P}) = (\wedge \text{PRED}) = \text{'P'} \\ @(\text{GLUE-PROPERNOUN P}).$$

In `glue-basic.lfg`, the argument `P` of `PROPERNOUN` is used to construct the f-structure semantic form as well as appearing as the value of the `MR` attribute in the AVM meaning constructor. If it is desirable for the f-structure `PRED` value to be different from the `MR` value of the AVM meaning constructor, the `PROPERNOUN`

⁵General documentation of the use of templates in XLE is available here: <https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/walkthrough.html#W.templates>

template would have to be defined to take two arguments, one providing the PRED value and the other providing the MR value.

GLUE-PROPERNOUN simply calls GLUE-RELO-MC (mnemonic for “meaning constructor for relation of arity 0”: in other words, a meaning constructor that requires no arguments). It specifies the first and second arguments of the template as $\hat{}$ and e for all proper nouns, and passes in the value of P as the third argument.

$$(14) \quad \text{GLUE-PROPERNOUN}(P) = @(\text{GLUE-RELO-MC } \hat{} \ e \ P).$$

In the `glue-basic.lfg` grammar, it would also have been possible for the PROPERNOUN template to call GLUE-RELO-MC directly, providing the arguments $\hat{}$ and e . The intermediate template GLUE-PROPERNOUN allows for the possibility that in scaling up to a more complete grammar, additional semantic specifications may be associated with the GLUE-PROPERNOUN template, besides the definition of the meaning constructor.

The definition of GLUE-RELO-MC is:

$$(15) \quad \begin{aligned} \text{GLUE-RELO-MC}(\text{FSTR } \text{TY } P) = & @(\text{GLUE-SEMSTR } \text{FSTR } \%sstr \ \text{TY}) \\ & @(\text{GLUE-MEANING } \%sstr \ P) \\ & \%sstr \ \$ \ (\text{FSTR } \text{GLUE}). \end{aligned}$$

This template calls the two basic templates GLUE-SEMSTR and GLUE-MEANING. The call to GLUE-SEMSTR specifies properties of the AVM $\%sstr$: it has an attribute SEMSTR whose value is FSTR, and it has an attribute TYPE whose value is TY. The call to GLUE-MEANING provides the value P for the attribute MR in $\%sstr$. The final line requires the f-structure $\%sstr$ to appear as a member of the GLUE set in the f-structure FSTR.

When the template GLUE-RELO-MC is called with arguments $\hat{}$, e , and Kim , an AVM $\%sstr$ is created which corresponds to the simple meaning constructor $\text{Kim}:\uparrow_e$. This AVM has three attributes: SEMSTR, whose value is $\hat{}$; TYPE, whose value is e ; and MR, whose value is Kim . The final line of this template specifies that $\%sstr$ is a member of the GLUE set in the f-structure FSTR. Thus, the template call $@(\text{PROPERNOUN } \text{Kim})$ produces the f-description given in (3).

2.4.3 Meaning constructors requiring arguments: Intransitive verbs

In `glue-basic.lfg`, the lexical entry for the intransitive verb *smiled* is:

$$(16) \quad \text{smiled} \quad V \quad * \quad @(\text{VERB-SUBJ } \text{smile}) \\ @VPAST.$$

The template VPAST specifies a past tense feature in the f-structure; we do not discuss this template here. The template VERB-SUBJ is defined as:

$$(17) \quad \text{VERB-SUBJ}(P) = (\hat{} \ \text{PRED}) = 'P < (\hat{} \ \text{SUBJ}) >' \\ @(\text{GLUE-VERB-SUBJ } P).$$

As with the proper name template described in the previous section, the template argument P is used to define both the f-structure semantic form and the MR value of the AVM meaning constructor. If this is not desirable, the template VERB-SUBJ should be defined to take two arguments, one specifying the

semantic form and the other the value of the MR feature in the AVM meaning constructor.

The template GLUE-VERB-SUBJ is defined as:

$$(18) \quad \text{GLUE-VERB-SUBJ}(P) = @(\text{GLUE-REL1-MC } (\wedge \text{ SUBJ}) \text{ e } \wedge \text{ t } P).$$

As with the GLUE-PROPERNOUN template, the GLUE-VERB-SUBJ template simply calls GLUE-REL1-MC (mnemonic for “meaning constructor for relation of arity 1”: in other words, a meaning constructor that requires one argument). In scaling up to a more complete grammar, there may be additional semantic specifications associated with GLUE-VERB-SUBJ.

The template GLUE-REL1-MC is defined as:

$$(19) \quad \begin{aligned} \text{GLUE-REL1-MC}(A1 \ A1TY \ FSTR \ TY \ P) = & @(\text{GLUE-SEMSTR } FSTR \ \%sstr \ TY) \\ & @(\text{GLUE-SEMSTR } A1 \ (\%sstr \ ARG1) \ A1TY) \\ & @(\text{GLUE-MEANING } \%sstr \ P) \\ & \%sstr \ \$ \ (FSTR \ GLUE). \end{aligned}$$

The first, third, and fourth lines of this template are the same as for the template GLUE-REL0-MC: they specify that the meaning constructor in the GLUE set of this verb is called `%sstr`, that it has an attribute SEMSTR with value FSTR, and that it has an attribute TYPE with value TY. The additional specification in the second line adds an attribute ARG1 to the structure, whose value for the SEMSTR feature is A1, and whose value for the TYPE feature is A1TY. When the template GLUE-REL1-MC is called with arguments $(\wedge \text{ SUBJ})$, `e`, \wedge , `t`, and `smile`, the resulting f-description is as in (5).

2.4.4 Scope of local names

It is important to be aware of the scope of local names (variables prefixed with `%`) in XLE – it is limited to the c-structure category in which the variable is used.⁶ This means that every time a given local name (for example: `%test`) is used within the same c-structure category, it refers to the same object – if there are two template calls using the same local name (`%test`) within one lexical entry, these template calls will impose constraints on the same object (one that corresponds to `%test`).

Depending on the intended effect, such behaviour of local names with respect to their scope may be a feature (when it is the intention to constrain the same object by separate template calls) or it may be undesired (when the intention is to impose constraints on two distinct objects by separate template calls) – this is why it is crucial to be aware of this when using local names.

This practical issue arises in sample grammars when a given c-structure category contributes more than one glue premise. For instance, as explained in the sample grammar `glue-basic.lfg`, the template GLUE-NOUN0-MC providing the meaning constructor for (common) nouns uses the local name `%sstrn`, because it must be different from the local name `%sstr` used by the template GLUE-QUANT-MC – both templates are called by the template QUANT called in the lexical entry of the quantifier “everyone”. Another example can be found in the

⁶<https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/notations.html#N4.1.6>: “A local name can be used as a variable whose scope is limited to the schemata associated with a particular category or lexical item.”

template `GLUE-ADJO` which provides two meaning constructors for prenominal adjectives – the call to the template `GLUE-REL1-MC` uses the local name `%sstr` to provide the meaning constructor for the basic meaning of the adjective, while the call to the template `GLUE-ADJ-MODIFIER` provides the meaning constructor combining the adjective with the noun by calling the template `GLUE-MODIFIER1` which uses the local name `%sstrm` to build this meaning constructor.

3 Sample grammars

The repository contains several XLE+Glue sample grammars. These build on the templates discussed in Section 2.4, and they also provide templates for further basic constructions, in particular, adjectival modifiers and verbs taking sentential complements. Each sample grammar is documented by means of comments describing its rules, lexical entries, and templates. The sample grammars that are currently available in the repository are:

`glue-basic.lfg` Corresponds closely to the discussion in this manual. `GLUE` attributes appear at f-structure, and meaning constructors refer to f-structures.

`glue-basic-semstr.lfg` Similar to `glue-basic.lfg`, but `GLUE` attributes appear at semantic structure and refer to semantic structures.

`glue-basic-semparser.lfg` Similar to `glue-basic.lfg` in that `GLUE` attributes appear at f-structure. Meaning representations are typed formulas of the lambda calculus, as described in Section 5.3. When parsing with this grammar, please follow the instructions in Section 5.3 to modify the `xlerc` file to activate the semantic parser.

`glue-basic-flat-encoding.lfg` Similar to `glue-basic.lfg`, but instead of encoding meaning constructors as AVMS, it uses the alternative string-based, flat encoding described in Section 7.1.

4 Formatting for glue premises: General representational issues

4.1 Using CONCAT

XLE provides the built-in template `CONCAT`⁷ which is used to concatenate an arbitrary number of elements – `CONCAT` joins all its arguments except the last one which is a local name (variable) containing the result of `CONCAT`. For instance: `@(CONCAT a b c %OUTPUT)` results in `abc` as the value of the local name `%OUTPUT`.

To avoid potential issues, it is safest to pass all special characters listed in section 4.2 as separate arguments to `CONCAT`.⁸

⁷<https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/notations.html#CONCAT>

⁸Minimally, it seems the following must be passed as separate arguments to `CONCAT`:

- every underscore (`_`)

4.2 Special characters in XLE

The following characters need to be escaped in XLE using a backquote (‘):⁹

- bound: < > () { } []
- oper: * + - & | \ / ~ ^ \$ ‘ # ’
- space: [space] \t [tab] \r [carriage return]
- mark: , : _ ? = ! % @ . ;

4.3 Using backquote-region.el

`backquote-region.el` is an emacs module aimed at making it easier to format strings for use with XLE+Glue. It provides two commands:

- `backquote-region` escapes characters listed in section 4.2 with a backquote (‘), so that the resulting string can be used as the value of the attribute MR.
- `backquote-space-region` escapes the same characters as `backquote-region`, but it also surrounds them with whitespace, so that the resulting string can be passed to `CONCAT` template without issues (see section 4.1).

To use `backquote-region.el`, put it in a directory (for instance: `~/elisp`), and add the following two lines to your `.emacs`:

```
(add-to-list 'load-path "~/elisp")
(load-library "backquote-region")
```

5 Semantic representations and the prover

5.1 Parsing linear logic formulas

The Glue semantics workbench encodes linear logic formulas in a simple string format that is visually similar to the actual symbols used in linear logic. For example, the \multimap symbol is replaced by `-o`. Examples (20) and (21) describe the formation of linear logic formulas that are well-formed from the perspective of the parser implemented in the prover. Most importantly, the parentheses around linear logic formulas are obligatory. Type declarations should either be applied to all constants and variables, or to none. Type declarations are indicated by an underscore, e.g., `_t`. Universal quantification over linear logic formulas is encoded in terms of an upper-case *A* followed by some variable (e.g. *X*) and a dot. As of now, the scope of a linear logic quantifier is the rest of the formula and does (and should) not need to be indicated via parentheses or brackets.¹⁰ This means, linear logic quantifiers behave differently from the first-order logic quantifiers introduced in the next section.

• [when it is the first character of the first argument of `CONCAT`

⁹Source: John Maxwell, PC.

¹⁰This behaviour is currently investigated and might change in the future.

- (20) **Atomic elements:**
 - a. *Constants:* String of lower-case alphanumeric characters; optionally with type declaration
 - b. *Variables:* String of upper-case alphanumeric characters; optionally with type declaration
- (21) Linear logic formulas:
 - a. *Linear implication:* $(\dots \multimap \dots)$, where \dots represents any well-formed formula
 - b. *Linear Quantification:* $AX. \dots$, where \dots represents any well-formed formula
- (22) **Available types:** e, s, v, t

5.2 System without the semantic parser

The XLE+Glue system has two modes of representation, with and without the semantic parser. In the mode of operation that does not make use of the semantic parser, any (string-based) format of semantic representation is compatible with XLE+Glue. This is the mode used by the demo grammars `glue-basic.lfg` and `glue-basic-semstr.lfg`, which include semantic representations that are treated as unanalysed strings by the workbench. In this simple format, functional application is expressed in the output by wrapping the argument in parentheses and concatenating functor and argument.

- (23) Combining $(1 \multimap 0)$: smile and 1 : Kim
to: 0 : smile(Kim)

5.3 Using the semantic parser

The Glue semantics workbench provides a semantic parser for lambda expressions supporting alpha- and beta-conversion. This section describes how to activate semantic parsing for your grammar and the guidelines for writing lambda expressions that can be parsed by the semantic parser. Integrating the semantic parser into your grammar is simple: In your `xlerc` file change the value for the variable `semParser` to 1 instead of 0. **Warning:** Once activated, unparsable input on the meaning side will result in a parsing error. Furthermore, the current version of the semantic parser is completely independent of the glue side, which means that type restrictions need to be manually added. Furthermore, eta-conversion is not possible. This may change in the future.

The semantic parser parses lambda expressions and first-order logic terms. First-order predicates are encoded in the classic prefix notation. There is no convention with respect to the casing of predicates or constants. Furthermore, constants and predicates are not themselves typed.

- (24) `like(mary, semantics)`

5.3.1 Lambda expressions

Lambda expressions are introduced via a scope defining bracket and a slash, followed by the variable that the lambda operator binds. Variables require a

type declaration to be distinguished from constants. This is done by using an underscore and one of the following types.¹¹

(25) **Available types:** e, s, v, t

Bound occurrences of a variable should not be typed again. The scope of the lambda function is separated from the binder via a dot. It can be any kind of well-formed (lambda) expression.

(26) a. `[/x_e.sleep(x)]`
 b. `[/x_e.[/w_s.sleep(x,w)]]`
 c. `[/P_<e,t>.[/Q_<e,t>.[/x_e.(P(x) & Q(x))]]]`

5.3.2 Complex types

Complex types consist of atomic types, commas and angular brackets, e.g.:

(27) a. `< e, t >`
 b. `<< e, t >, t >`
 c. `< e, < s, t >>`

5.3.3 Logic operators

The basic logic operators \wedge, \vee and \rightarrow can be used as infix operators (see (28)a - (28)c), although their scope has to be defined via brackets or parentheses. Brackets can indicate operator scope and quantifier scope simultaneously (see (28)d). Other operators must be encoded as first-order logic predicates, i.e. in prefix notation (see (28)e).

(28) a. **Logical ‘and’ (&):** `(P(x) & Q(x))`
 b. **Logical ‘or’(v):** `(P(x) v Q(x))`
 c. **Logical ‘implication’(->):** `(P(x) -> Q(x))`
 d. **Variant with brackets:** `Ex_e[P(x) & Q(x)]`
 e. **Prefix notation:** `equals(x,y)`

5.3.4 Quantifiers

Quantifiers are introduced via the upper case letters **A** and **E**. Their scope variable is introduced with an appropriate typing. The scope is defined via brackets.

(29) a. `Ex_e[dog(x) & bark(x)]`
 b. `Ax_e[cat(x) -> sleep(x)]`

5.3.5 Functional application

Functional application steps such as the lambda terms for quantifiers are determined contextually, e.g., $P(x)$ and $Q(x)$ in the example below. More concretely, the P and Q variables over predicates followed by the x variable as an argument are translated as functional application steps (apply P/Q to x), rather than as a one-place predicate with a bound variable ($P(x)$).

¹¹Contact Mark-Matthias Zymla if you require support for a new type.

$$(30) \quad [P_{-} \langle e, t \rangle . [Q_{-} \langle e, t \rangle . \text{Ex_e}[P(x) \ \& \ Q(x)]]]$$

5.3.6 Lexical entries in XLE

We briefly present a possible way of encoding appropriate semantic representations in XLE. As pointed out in section 4.1, the challenge is to overcome the issue of escaping certain characters while preserving variable unification. One way to achieve this is to use the @CONCAT template to separate variables and string parts of a meaning representation.

$$(31) \quad \begin{aligned} \text{GLUE-REL1-MC(A1 A1TY FSTR TY P)} = & \quad @(\text{GLUE-SEMSTR FSTR \%sstr TY}) \\ & @(\text{GLUE-SEMSTR A1 (\%sstr ARG1) A1TY}) \\ & @(\text{CONCAT ' ['/x A1TY ' . P '(x') '] \%MEANING}) \\ & @(\text{GLUE-MEANING \%sstr \%MEANING}) \\ & \%sstr \$ (\text{FSTR GLUE}). \end{aligned}$$

This approach is still being tested. Further feedback for encoding meaning representations according to the specifications given above is welcome.

6 Prolog rewrite component

The Prolog rewrite component takes the Prolog output of an XLE parse as input.¹² It does not rely on any particular assumptions about where the GLUE attributes must appear; as discussed in Section 3, GLUE attributes and their values are a part of f-structure in our sample grammar `glue-basic.lfg`, while our sample grammar `glue-basic-semstr.lfg` places them at s-structure. Indeed, the system works even if some GLUE attributes appear at f-structure, and others appear in other structures. It is also not necessary for the meaning constructors to be distributed in any particular way in the structure in which they appear; the system simply gathers up all members of every GLUE set in the input representation, rewrites them into the standard format, and passes the resulting set of standard-format meaning constructors to the prover. Thus, decisions about which structure hosts the GLUE attribute and its values should be made on the basis of linguistic considerations, and are not determined by properties of the implementation. Attributes other than the GLUE attributes and their values are ignored and discarded by the rewrite component.

Each element of the GLUE set provides one premise: as explained above, the value of MR provides the semantic side, while SEMSTR, TYPE, and the ARG1...N attributes provide the glue side.¹³

The SEMSTR labels are instantiated to the numeric labels provided by XLE. Because the numeric indexing for semantic forms in the Prolog output format is independent of the numeric indexing for other structures (for example, there may be an f-structure with numeric index 1 and also a semantic form with numeric index 1 in the same f-structure), the numeric index of a semantic form is additionally prefixed with an S, e.g., S1, to ensure uniqueness of indices. As

¹²https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle.html#Prolog_Output

¹³In fact, only the attributes MR, SEMSTR, TYPE, and FORALL have a special status. All other attributes are assumed to represent arguments, which are consumed according to alphabetical order. It would also be possible to use A, B, C; A1, A2, A3; or any other alphabetically ordered series of attributes for arguments. It is not possible to substitute other names for the special attributes MR, SEMSTR, TYPE, FORALL.

described above, the `FORALL` attribute `F` is used as reference to linear quantification. Different linear quantifiers in a computation are distinguished by combining the label `F` with its unique f-structure index.

Example (32-a) shows the output produced by the rewrite component for a generalized quantifier encoded in AVM format as (32-c) (as discussed in example (8)), corresponding to the standard format meaning constructor in (32-b). All of the conventions discussed above are illustrated in (32-a): `F11` is bound by a universal quantifier, `s1` refers to the semantic form whose index is 1, and 4 refers to the f-structure whose index is 4.

- (32) a. `AF11.t.((s1.e -o s1.t) -o ((4.e -o F11.t) -o F11.t))`
 b. $\forall F11_t.((s1_e \multimap s1_t) \multimap ((4_e \multimap F11_t) \multimap F11_t))$

$$\text{c. } 4: \left[\begin{array}{l} \text{PRED } 1 \\ \text{GLUE } \left\{ \begin{array}{l} \text{MR } \textit{every} \\ \text{ARG1 } \left[\begin{array}{l} \text{ARG1 } \left[\begin{array}{l} \text{SEMSTR } 1 \\ \text{TYPE } e \end{array} \right] \\ \text{SEMSTR } 1 \\ \text{TYPE } t \end{array} \right] \\ \text{ARG2 } \left[\begin{array}{l} \text{ARG1 } \left[\begin{array}{l} \text{SEMSTR } 4 \\ \text{TYPE } e \end{array} \right] \\ \text{SEMSTR } F11 \\ \text{TYPE } t \end{array} \right] \\ \text{FORALL } F11 \\ \text{SEMSTR } F11 \\ \text{TYPE } t \end{array} \right\} \end{array} \right]$$

7 Illustrating the flexibility of the system

7.1 Meaning constructors encoded as strings

Section 2 describes the f-structure encoding of meaning constructors as AVMs, using attributes such as `MR`, `ARG1`, `SEMSTR` and `TYPE`. In this encoding, the embedding in the AVM representation reflects the structure of the linear logic expression that is encoded, since material on the left-hand side of a linear implication is represented as the f-structure value of an attribute such as `ARG1`.

This section describes an alternative encoding: the string-based, flat encoding. In this encoding, the substrings of the meaning constructor are encoded as values of the attributes in a single AVM, which are concatenated together to produce the input to the prover. This is in some ways a simpler encoding, since it does not require the construction of a complex AVM to reflect the structure of the linear logic term. However, it requires a detailed understanding of the input format required by the GSWB prover, and it is also easier to make mistakes in the encoding, which can make it harder to use.

The flat encoding is illustrated in (33), where 1 is the label assigned by XLE to the outermost f-structure:

- (33) a. `Kim:1.e`

$$\begin{array}{lcl}
\text{b.} & 1 : & \left[\begin{array}{cc} \text{PRED} & \text{'KIM'} \\ & \left\{ \begin{array}{cc} \text{A1} & \text{KIM} \\ \text{A2} & : \\ \text{A3} & 1 \\ \text{A4} & - \\ \text{A5} & e \end{array} \right\} \end{array} \right] \\
\text{c.} & \begin{array}{ccccc} \text{Kim} & : & 1 & - & e \\ \text{a1} & \text{a2} & \text{a3} & \text{a4} & \text{a5} \end{array} & &
\end{array}$$

In this encoding, the substrings of the meaning constructor input to the GSWB prover are encoded directly. By convention, this encoding uses the attributes $A1, A2, \dots$, but in fact any attributes can be used except for the special attribute MR , which indicates to the transfer component that the standard encoding is being used. It is possible to have meaning constructors encoded in the string-based format and the standard embedded format coexisting in the same f-structure, or even in the same GLUE set.

In the transfer component using the flat encoding, the attributes are sorted into alphabetical order, and the substrings of the meaning constructor are concatenated according to that ordering. (33-c) shows the correspondence between the attributes of the f-structure in (33-b) and the resulting meaning constructor. A cautionary note: if there 10 or more meaning constructor substrings encoded via attributes $a1 \dots a10 \dots$ in an AVM, the attribute $a10$ will sort alphabetically between the attributes $a1$ and $a2$; in that case, therefore, you should instead prefix the single-digit attributes with 0 ($a01, a02, \dots a10, a11, \dots$) to ensure that the values of the attributes are concatenated in the correct order.

The lexical entry for *Kim* in the sample grammar `glue-basic-flat-encoding.lfg` calls the `PROPERNOUN` template as shown in (12), and `PROPERNOUN` is defined in terms of the template `GLUE-RELO-MC` as shown in (13). In the string-based, flat encoding, `GLUE-RELO-MC` is defined as follows:

$$\begin{array}{lcl}
(34) & \text{GLUE-RELO-MC(FSTR TY P)} = & \begin{array}{l} (\%sstr A1) = P \\ (\%sstr A2) = ' : ' \\ (\%sstr A3) = FSTR \\ (\%sstr A4) = - \\ (\%sstr A5) = TY \\ \%sstr \$ (FSTR GLUE) . \end{array}
\end{array}$$

The value of the attribute $a1$ is the meaning term, which is the first component of the meaning constructor for *Kim*. The value of $a2$ is the colon separating the meaning side of the meaning constructor from the glue side, which must be quoted with a backquote. The value of $a3$ is the f-structure for *Kim*, the value of $a4$ is the underscore separating the f-structure from its type, and the value of $a5$ is its type, as specified by TY in the template call to `GLUE-RELO-MC`. Further examples can be found in the sample grammar `glue-basic-flat-encoding.lfg`.

8 Setting up and running the system

For full details on setting up and running the XLE+Glue system, please consult:

<https://github.com/Mmaz1988/xle-glueworkbench-interface/blob/master/README.md>

The basic steps are:

- Set up XLE according to these instructions:
<https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle.html#SEC1.5>
- Download the XLE+Glue system and sample grammars.
- Navigate to the XLE+Glue folder that you downloaded and run XLE. The `xlerc` file that comes with the system automatically loads the sample grammar `glue-basic.lfg`, but you can edit it to load another sample grammar or your own grammar.
- The `xlerc` file sets up the new functionalities. It requires on the directory `src`, which is part of the github repository. It is easiest, to just keep the structure of the github repository to run the system.
- You will find an additional entry “Semantics” in the f-structure (or s-structure) “Commands” menu. Choose the “Semantics” command to start up the prover and produce a proof from the meaning constructors contributed by the sentence you parsed.

A Appendix: Running XLE in Windows 10 Subsystem for Linux

We have found that using the Windows Subsystem for Linux is an easy way of running the Linux version of XLE in Windows 10. Here is an overview checklist for this means of running Linux in Windows 10 (last updated: May 2020).

1. Follow the instructions on this webpage for installing a Linux distribution in Windows 10:

<https://docs.microsoft.com/en-us/windows/wsl/install-win10>

(Note that your Windows drives can be accessed from within Linux via the Linux directory `/mnt`: for example, the C drive can be accessed as `/mnt/c`.)

2. Install the Xming X-server to allow display of graphics, including XLE windows.
 - (a) Download Xming here: <http://www.straightrunning.com/XmingNotes/>
Be sure to install both Xming and Xming-fonts.
 - (b) Run Xlaunch to create and save your profile. There is no need to change the default values. Save the profile to the directory where the XLaunch application is located by clicking on “Save configuration” at the end of the XLaunch dialogue and navigating to the Xming directory.
 - (c) Add “`export DISPLAY=:0.0`” to the end of your Linux .profile file if you are using the Bash shell. (If you are not using the Bash shell, try “`setenv DISPLAY 0.0`” in your .login file.)
 - (d) Depending on how you have set up Xming, you may have to launch Xming whenever you start up a Linux process.

3. You may find it convenient to allow copy/paste to and from Linux windows. You can get this by right-clicking on the white bar at the top of a Linux window and choosing “Properties”, and making sure that the check box next to “Use Ctrl+Shift+C/V as Copy/Paste” is ticked.
4. Install the Linux version of XLE and the XLE+Glue system according to the instructions in Section 8.
5. To run the XLE+Glue system, you will also need to install SWI-Prolog and Java on your Linux system. For Ubuntu, the following works:
 - (a) SWI-Prolog:


```
sudo apt-add-repository ppa:swi-prolog/stable
sudo apt-get update
sudo apt-get install swi-prolog
```
 - (b) Java:


```
sudo apt install default-jdk
```

References

- Crouch, Dick, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell III, and Paula Newman. 2017. *XLE Documentation*. Palo Alto Research Center.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar*, volume 34. New York: Academic Press.
- Dalrymple, Mary, John Lamping, and Vijay A. Saraswat. 1993. LFG semantics via constraints. In *Proceedings of the 6th Meeting of the EACL*, Pages 97–105. European Association for Computational Linguistics, Utrecht.
- Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.
- Meßmer, Moritz and Mark-Matthias Zymla. 2018. The Glue Semantics Workbench: A Modular Toolkit for Exploring Linear Logic and Glue Semantics. In M. Butt and T. H. King, editors., *Proceedings of the LFG’18 Conference, University of Vienna*, Pages 249–263. Stanford, CA: CSLI Publications.