The LOGON MT infrastructure

Ling 567
February 24, 2015

Overview

- MRS and MT: Some history

- The Grammar Matrix and massively multilingual MT

- The LOGON architecture

  - Processing steps

  - Transfer rules

  - VPM

- Lab 8 practicalities

- Next week: Transfer rules

MRS and MT: Some history

- Copestake et al 1995: Original motivation for MRS included MT applications

- Resolving scope ambiguities is hard, and usually not necessary

  - Logical form equivalence is undecidable even in FOPL (Shieber 1993)

- Mimicking syntactic structure in semantics makes transfer harder

  - *fierce black cat <> gato negro y feroz* (Spanish)

  - *young black bull <> novillo negro*

- MRS gives logical forms with less syntactic complexity and underspecification wherever possible.

MRS and MT: Some history

- MRS originally developed in the context of Verb*Mobil* but not fully deployed for transfer-based MT in that project.

- In 2003, LOGON (Oepen et al 2004, 2007) up the thread and builds the first MRS-based MT system. (Norwegian -> English; tourism brochures)

- Input is LFG, with MRSs projected from f-structure.

- Output is generated by the English Resource Grammar (HPSG; Flickinger 2000)
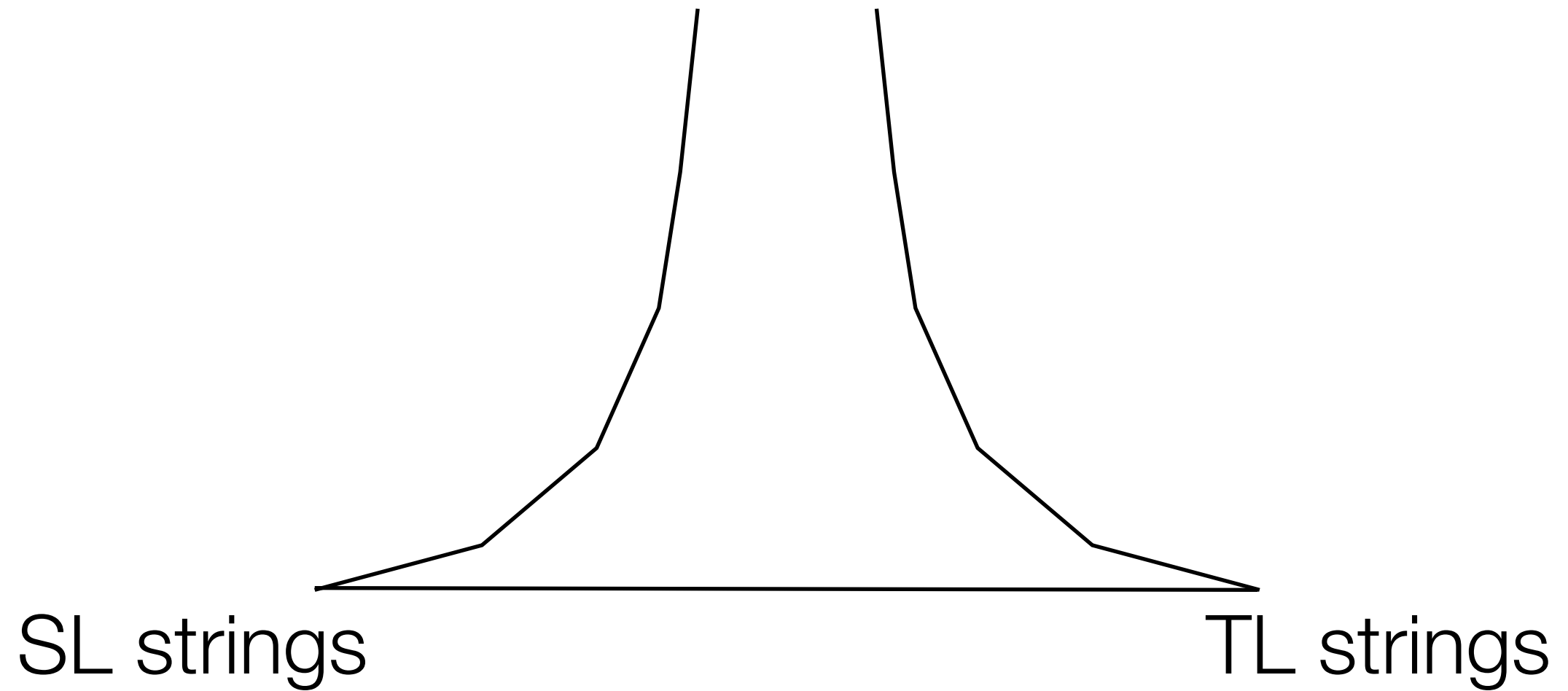
Vauquois Pyramid (ObMT Triangle)

Interlingua

SL strings   TL strings
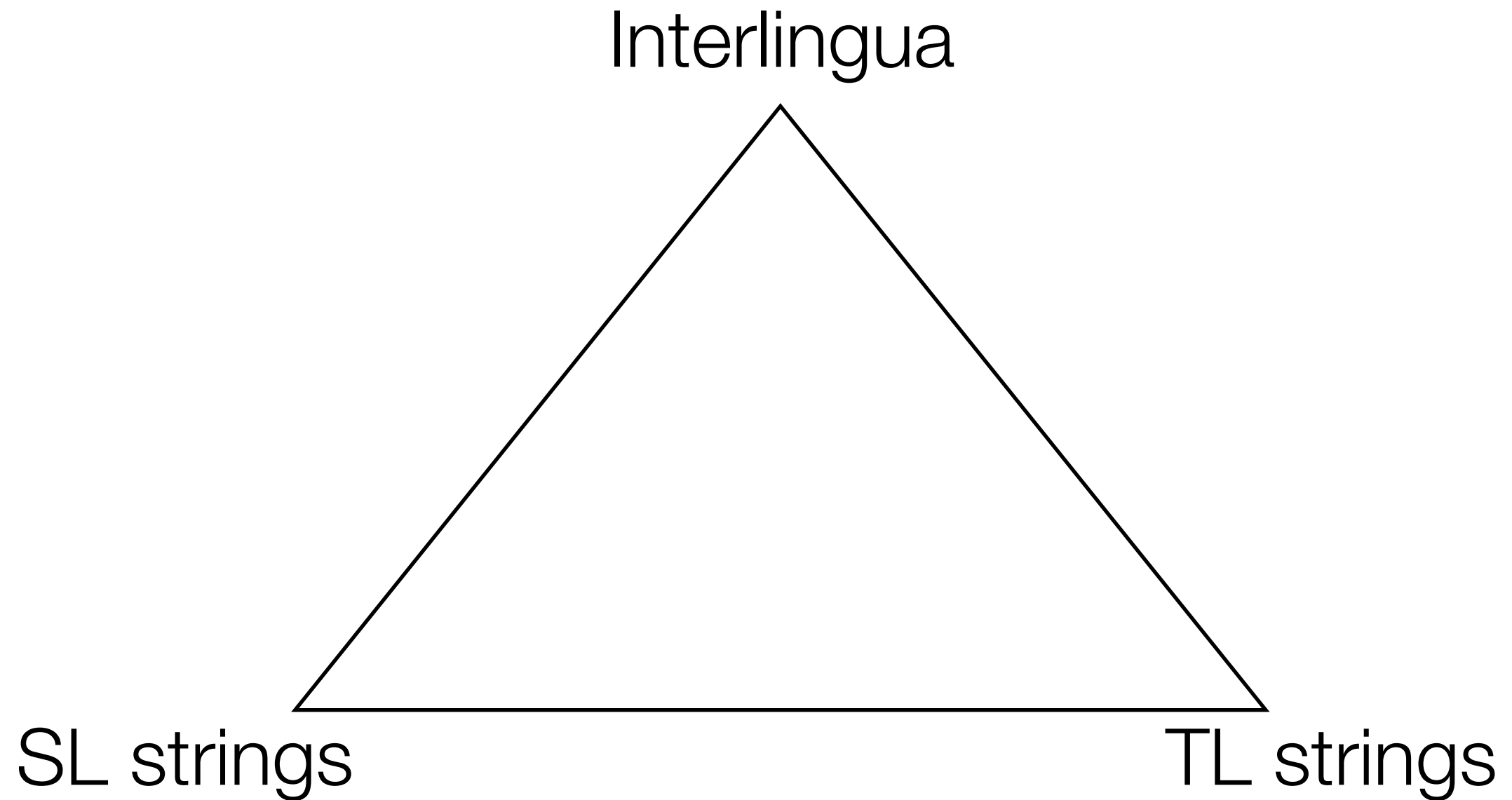
Is MRS an interlingua?

- Could MRS be used to encode an interlingua?

- Could our grammar produce such an MRS-encoded interlingua?

Copestake Volcano

SL strings   TL strings

Massively Multilingual MT

- Problem of combinatory explosion (*n x n*):

- 2 languages: 2 sets of transfer rules

- 4 languages: 9 sets of transfer rules

- 24 languages: 552 sets of transfer rules

- 6000 languages: 35,994,000 sets of transfer rules

What are the alternatives?

- Design an interlingua (or select a pivot language), and create two grammars for each language

  - strings <> ordinary MRS

  - ordinary MRS <> interlingua (transfer grammar)

- Hybrid interlingual/transfer-based model

  - partial lexical interlingua or PanDictionary-derived rules

  - TL-side "accommodation" transfer grammars: O(n)

  - transfer matrix to capture generalizations

- How far will approach 2 scale?

- How much mismatch is there?

Mismatch: Translation divergences (Dorr 1994)

- Categorial divergence: Translation of words in one language into words that have a different part of speech in another language.

- Conflational divergence: The translation of two or more words in one language into one word in another language

- Structural divergence: The realization of verb arguments in different syntactic configurations in different languages.

- Head swapping divergence: The inversion of the structural dominance relation between two semantically equivalent words when translating from one language to another.

- Thematic divergence: The realization of verb arguments in different configurations that reflect different thematic to syntactic mapping orders.

MRS 'harmonization' helps

- Just because it's not an interlingua doesn't mean the grammars can't be brought closer together.

- Example 1: Demonstratives (adjectives v. determiners)

- Example 2: COG-ST et al, reduction in quantifier-rel inventory

- Further potential for harmonization: pronouns v. pro-drop (but cf. information structure marking on overt pronouns)

- Other examples?

LOGON processing steps

- Parse in source language

  - visualization tools for parses and MRSs

- Apply source language's transfer grammar to produce new MRS

  - visualization tools for transfer outputs

- Generate in target language from new MRSs

  - visualization tools for input MRSs

  - compare to MRS produced by parsing expected output

  - generator chart

Anatomy of a transfer rule

- Quadruple: [CONTEXT:] INPUT [!FILTER] -> OUTPUT

- Each item above is a (partial) MRS

- Rules apply to complete MRSs to produce partially rewritten MRSs.

- Resource sensitive: INPUT is consumed in producing OUTPUT.

- CONTEXT: Additional properties beyond the INPUT that must be satisfied. (Not consumed.)

- FILTER: Negative constraints; contexts in which the rule should not apply.

Anatomy of a transfer rule

- Rules can be obligatory or optional.

- Optional rules produce non-determinism in the transfer process.

- Pairing each optional rule with one obligatory rule cuts down the transfer search space.

- Rules can also be grouped into sets for 'extrinsic' ordering (which we probabl won't need).

- Handled with chart-based processing.

Types and translation

- Many transfer rules share most of their properties, differing only in lexical predicates/other small details.

  ‣ Define types of transfer rules, with particular instances, analogous to lexical types and lexical entries.

- Types mentioned in transfer rules will unify with compatible types in actual MRS.

- In addition, the generator will allow some unification of different (but compatible) types for feature values.

Example type

```
monotonic_mtr := mrs_transfer_rule &
[ CONTEXT.HOOK.LTOP #h,
  INPUT.HOOK.LTOP #h,
  OUTPUT.HOOK.LTOP #h ].
```

Example rule instance

```
pro-insert-arg1-mtr := monotonic_mtr &
[ INPUT.RELS <! !>,
  CONTEXT.RELS <! [ ARG0.SF prop-or-ques,
                    ARG1 #x & x ] !>,
  FILTER.RELS <! [ ARG0 #x ] !>,
  OUTPUT [ RELS <! [ PRED "_pronoun_n_rel",
                     ARG0 #x,
                     LBL #larg ],
                   [ PRED "exist_q_rel",
                     ARG0 #x,
                     RSTR #harg ] !>,
         HCONS <! qeq &
                  [ HARG #harg,
                    LARG #larg ] !> ],
  FLAGS.EQUAL < #x > ].
```

What about features of indices?

- Can't change value from input to output while maintaining identity of index with other positions.

- Person and number can be harmonized (in principle at least) by extending hierarchies on both sides, but we can't harmonize between PERNUM and separate PER and NUM features.

- Tense and aspect (and others) can likewise be harmonized at least somewhat, but inventories vary greatly.

- Variable property mapping allows grammar-internal variable properties to differ from grammar-external universe.

  - We'll use this for harmonization (e.g., of PERNUM) and setting of defaults.

A side note on gender

- Represented in MRS because of its role in reference resolution.

- Pretty language specific.

- You might think you want to keep it on pronouns and discard it on nouns, but even that only works for closely related languages.

- Long term solution: Anaphora resolution on the SL language side and assignment of gender properties to pronouns based on projections of this information.

- For now: drop gender through vpm.

```
;;; -*- Mode: TDL; Coding: utf-8 -*-
; A basic VPM for Matrix grammars.COG-ST : COG-ST
event              <> e            type-id <> type-id
ref-ind            <> x            uniq-id <> uniq-id
individual         <> i            familiar <> familiar
handle             <> h            activated <> activated
non_event          <> p            in-foc <> in-foc
*                  >> u            activ+fam <> activ+fam
semarg             << u            uniq+fam <> uniq+fam
                                   activ-or-more <> activ-or-more
SORT : SORT                        uniq-or-less <> uniq-or-less
   * <> *                          uniq+fam+act <> uniq+fam+act
   semsort << *                    fam-or-more <> fam-or-more
                                   fam-or-less <> fam-or-less
                                   uniq-or-more <> uniq-or-more
E.TENSE : E.TENSE                  activ-or-less <> activ-or-less
   * <> *


SF : SF
   prop <> prop
   ques <> ques                    PNG.PER : PER
   prop-or-ques >> prop-or-ques    1st <> 1st
   prop << prop-or-ques            2nd <> 2nd
   comm <> comm                    3rd <> 3rd
                                   * <> !
```

```
PNG.NUM : NUM                          E.MOOD : MOOD
  sg <> sg                               irrealis <> irrealis
  pl <> pl                               resemblative <> resemblative
  du <> du                               quotative <> quotative
  dist <> dist                           apparitional <> apparitional
  coll <> coll                           iterative <> iterative
  * <> !                                 potential <> potential
                                         * <> !

PNG.GEND : GEND
  animate <> animate
  inanimate <> inanimate
  human <> human
  nonhuman <> nonhuman
  * <> !


E.ASPECT : ASPECT
  continuative <> continuative
  * <> !
```

```
PNG.PERNUM : PER NUM
   1singular <> 1st singular
   2singular <> 2nd singular
   3singular <> 3rd singular
   1plural <> 1st plural
```

Practicalities

- Get one more test corpus example working

- Grammar clean-up

  - Reduce number of strings generated per input to (ideally) those that are motivated.

  - Harmonize MRSs

  - Use VPM to set defaults for e.g., ASPECT

```
      * >> no-aspect
      no-aspect << [e]
```

- Attempt translation

- Work on VPM

Overview

- MRS and MT: Some history

- The Grammar Matrix and massively multilingual MT

- The LOGON architecture

  - Processing steps

  - Transfer rules

  - VPM

- Lab 8 practicalities

- Next week: Transfer rules