The Lexicon in MT

R. Lee Humphreys
University of Essex
Wivenhoe Park
Colchester, UK

SALT Club MT Workshop
UMIST
2-3 June 1990

1 Exclusions
Not much to say on derivational and inflectional morphology

2 Assumptions
- Rule-based vs Information Theoretic
- Rules vs Principles
- 3 Conceptual components: Analysis - Transfer - Synthesis
  - Analysis
    * = computation from strings to some appropriate underlying representation
    * part of general NLP programme
  - Synthesis
    * = computation from underlying representation to surface strings
    * part of general NLP programme
  - Transfer
    * = matching up elements in underlying representations (difference grammar)
    * specific to MT programme
    * but should be assimilable to general interlevel mapping
  —Generality of Interlevel Mapping
  —Sequential vs Simultaneous resolution of constraints
  —Constraints and Modularity and Redundancy
Comment Possible at multiple levels simultaneously e.g. we could say that [AP N] in English goes to [N AP] in French as well as specifying more generally that restrictive mods go to restrictive mods. Certain transfer generalisations may well be best characterised at a constituency level. Possibility of redundant transfer specification.

Comment Notion of transfer modularity questionable - depends on autonomy of underlying (interface) representations: but can ARG1 of a given predicate be interpreted entirely without reference to associated string sets?

3 General structure

- Grammar = Rules + Lexicon

  Comment No particular view taken on partitioning of work. Everything can be in the lexicon if we want.

- Analysis/Synthesis: Monolingual dictionaries - relation between strings and underlying representation

  atomic e.g. house $\Leftrightarrow$ house1

  Comment Integrity of the word and "Fido $\Leftrightarrow$ Fido" theory of meaning

  non-atomic eg. vp flattening

- Transfer: Multilingual dictionaries a relation on elements of underlying representation

  atomic e.g. house1 $\Leftrightarrow$ maison1

  non-atomic ideally, a homomorphism
  in practice, more complex maps usually required

  e.g. compounds $\Leftrightarrow$ phrases
  e.g. argument-changing verb pairs etc

  \[
  \tau(S_{sl}, S'_{tl}) = \tau_{Lex}(S_{sl}, S'_{tl}) \land \tau_{Tns}(S_{sl}, S'_{tl}) \land \tau_{Asp}(S_{sl}, S'_{tl}) \land \ldots
  \]
4 Sample Entry from Monolingual Lexicon

PHON

blow

HEAD

[MAJ V]

VFORM	FIN

AUX 

-

SUBCAT

[ SYN|LOC HEAD [MAJ SUBCAT { } ], NP1 ]

LEX

+ 

TENSE

past

PRED

blowup1

ARG1

1

Types and Hierarchies

5 Sample Entry from Bilingual Lexicon

PHON

blow

HEAD

[MAJ V]

VFORM	FIN

AUX 

-

SUBCAT

[ SYN|LOC HEAD [MAJ SUBCAT { } ], NP1 ]

LEX

+ 

TENSE

past

PRED

blowup1

ARG1

1

Types and Hierarchies

6 Monolingual Entries

blow¹ /blow/ vi, vt ..... 1 [VP2A,C] (with air, wind or it as subject), move along, flow as a current of air: It was ~ing hard, there was a strong wind. It was ~ing a gale / great guns, there was a violent gale.... 2[VP15A,B] (of the wind) cause to move .... 3 .... 12 (compounds from the v) ~dry vt [VP6A] ... ~fly n common meat fly.....

blow² /blow/ n blowing: (Give your nose a good ~, clear it thoroughly.

have/go for a ~, go outdoors for fresh air

blow³ /blow/ n 1 hard stroke (given with the hand, a stick etc): He struck his enemy a heavy~on the head, at one ~; at a (single) ~, in a single effort: I killed six flies at a single ~...... 2 shock; disaster: His wife's death was a great ~to him

blow⁴ /blow/ vi ..(chiefly in pp as) full blown roses, wide open OALD
7 Structure of an Entry

**Headword fields** superscripts usually indicate (OALD) homonyms. (Problem: Nominalisations having the same wordform as source predicate usually part of same entry; different wordform, then different entry.)

**Pronunciation fields** something for the speech people

**Inflection fields** something for everyone else

**Definition fields** varieties of word meaning

- classic dictionary = relation on dictionary elements
- NLP = e.g. operations on database
- MT = translation via multilingual dictionary i.e. a string set in target language
- MT/NLP = relation on lexical primitives

**Example fields** ideally, redundant; typically, catch all for pattern gaps

**Senses and Patterns** Is there a systematic relation?

- Different synonyms, different selectional restrictions - noun classes
  - 13 blow up (a)bridges etc. (FR: exploser/sauter)
  - (b) tyres (FR: gonfler)
  - (c) photographs (FR: agrandir)
  - (d) reputations, events, situations (FR: exagerer)
    (OALD)
  Bridges, tyres, photographs - concrete
  Reputations, events, situations - abstract

- Argument Realisation (patterns)
  1. Sense-preserving alternations:
     - one sense, same valency, same focus, different arg realisation
       [VP2C,E] *The hat blew off/into the pond*
     - one sense, same valency, different focus
       *The wind blew the door down*
       *The door was blown down (by the wind) (Passive)*

       *A drought followed*
       *There followed a drought* (Presentational *it*/*there*)

       **Comment** OALD notes this as a possible variation in the realisation of an intransitive pattern - but no note on its availability.

     - one sense, same valency, ?same focus, particle shift
       *The wind blew the door down*
       *The wind blew down the door*
- one sense, same valency, different focus, different Aktionssart

  He emptied the tank of petrol
  He emptied petrol from the tank
  (N.B. Issue of variation in durative modification)

2. Productive sense-changing alternations
   — Causative/Anticausative
     The wind blew off his hat
     FR: Le vent a fait s'envoler son chapeau
     His hat blew off
     FR: Son chapeau s'est envolé

   Comment OALD notes the ref. blew his whistle vs The whistle blew but does not assign the two forms different senses

- Resultative
  He blew bubbles
  FR: Il a fait des bulles

- Middle
  This shirt washes well

   Comment OALD pattern assimilates this to plain intransitive

3. Idiosyncratic sense-changing alternations

11 spend money recklessly or extravagantly: ~£10
   on a dinner with a girl friend
   (OALD)

Modification Need for specification of possible modifier types e.g. psychverbs exclude locationals

8 Multi-Word Units

  • Typically, represent completely frozen units (sayings), failures of compositionality (idioms) or lexical selectional constraints on environment (collocations)

  • Typically, appear within sense fields i.e. a particular mwu associated with a particular headword sense.

  • General translation strategy: associate tree/graph fragments with corresponding t.l. tree/graph fragments e.g.

    #1 blow the gaff ⇔ #2 vendre la mèche
    where τ(#1,#2)

Some issues:
Operators Mel'cuk style LFs: reducing the transfer coding effort Support and Headedness

\[(X \text{ attack } Y) \text{ vs } (X \text{ made an attack on } Y)\]
\[(X \text{ raise/lift the blockade})\]
\[\text{Oper}_{12} (\text{attack}) = \text{make}\]
\[\text{Term}_{12} (\text{blockade}) = \text{raise}\]

MWU Syntax: Well-formedness Need to ensure that MWUs denoted by well-formed trees/graphs.

MWU Semantics Transformational Deficiencies and Modifiability

- Discourse referential status of idiom elements
e.g. *The donkey kicked the bucket*
- Blocking type constructors / modification

MWU: Size What constitutes an idiom chunk?

EN: He gave me *a blow-by-blow account*
FR: Il ne m'a fait grâce d'aucun détail

Operators and Idiom units
Dative shift: handled pre-IS

Reanalysis *To blow the gaff on somebody* $\Leftrightarrow$ *Dénoncer qn.* Reanalyse as *denounce*?

denounce X (as Y) (to Z)
blow the gaff on X (as Y) (to Z)
make a denunciation of X (as Y) (to Z)