Recognizing polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Palmorf's morpholymical headaches

Eckhard Bick
Southern Denmark University
lineb@hum.au.dk

Program chains used in the Morpholympics

Distributed morphology

Why polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Local disambiguation 1

Local word class disambiguation in lemmas

Local word class disambiguation on clitics

Some problems in retrofixing an integrated tagger-disambiguator for multi-tagging
- Palmorf does some tokenization only after disambiguation.
- Most adjectives with potential nominal function are not tagged as nouns by the analyzer, but identified as np-heads by the syntactic grammar, and - if wanted - later marked with a secondary noun tag.
- The analyzer itself uses some post-mature tags (e.g. "PR FX -" or MM"") instead of 2 separate tag lines. The ambiguity is resolved after syntax by dependency propagation.
- Palmorf expects running text input, and its pattern matching filters can get confused by spaces before punctuation, non-standard quotes, non-adjacent apostrophes etc.
- In a tagger-disambiguator, derivation overgeneration across word classes is no problem, since it can be contextually resolved, so the analyzer as such is not optimized for precision in this field.

Distributed morphology

Program chains used in the Morpholympics

Distributed morphology

Why polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Palmorf's morpholymical headaches

Eckhard Bick
Southern Denmark University
lineb@hum.au.dk

Program chains used in the Morpholympics

Distributed morphology

Why polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Palmorf's morpholymical headaches

Eckhard Bick
Southern Denmark University
lineb@hum.au.dk

Program chains used in the Morpholympics

Distributed morphology

Why polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Program chains used in the Morpholympics

Distributed morphology

Why polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Palmorf's morpholymical headaches

Eckhard Bick
Southern Denmark University
lineb@hum.au.dk

Program chains used in the Morpholympics

Distributed morphology

Why polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Program chains used in the Morpholympics

Distributed morphology

Why polylexicals (MWE) ?
- Recognizing polyvalent "verbs", "prepositions" etc., etc. is essential for content-based disambiguation and syntax, since it makes context patterns less complex.
- Recognizing polyvalent "nominal" and verb incorporation (either coocurrence, error com-fond) helps semantic disambiguation and MT.
  (Verb incorporation is currently inactivated in PALAVRAS)
- Recognizing "nominal" and treating them as nouns.
  a) creates simpler context for disambiguation of other word classes
  b) allows NE-attachment, e.g. "sham" recognition by Christian name first part
  c) prevents lowercase name parts, why, why from recovering other word classes
  d) allows a meaningful analysis of non-segmented numbers and punctuation

Palmorf's morpholymical headaches

Eckhard Bick
Southern Denmark University
lineb@hum.au.dk

Program chains used in the Morpholympics

Distributed morphology
Local disambiguation 2

Local gender/number disambiguation in hyphenates
MAP \( /c134 M \) TARGET (<hyphen> M/F) (1C (ADJ M) OR (PCP M)) ; # comunista-adjunto, azul-claro
MAP \( /c134 S \) TARGET (<hyphen> S/P) (1C (ADJ M) OR (PCP S)) ;
MAP \( /c134 M \) TARGET (M/F) (-1C (N M) OR (ADJ M) OR (PCP M) LINK 0 <hyphen>) ;
MAP \( /c134 S \) TARGET (S/P) (-1C (N S) OR (ADJ S) OR (PCP S) LINK 0 <hyphen>) ;

Local mode disambiguation in hyphenates
REMOVE (IMP 2S) (0 <hyfen>) (1 PERS) (NOT 1 (2S ACC)) ; # incorpora-se, ikke: lava-te

Overgenerating affixation
REMOVE (<DERS <DERS) (0 PROP) ; REMOVE (<DERP <DERP) (0 PROP) ;
REMOVE (<DERS -aço [AU/PEJ] > <DERS -ão [AU]> ) ; # no double augmentatives
REMOVE (<DERS <DERS -ia [ABSTR]) ; # '-ia' not after other suffixes
REMOVE (<DERS <DERS -or [ABSTR]) ; # '-or' not after other suffixes

Local disambiguation after NE-type-recognition
REMOVE (<DERS 0 <DERS) (0 <exheur>) ; # after name recognition in samtrad
REMOVE (exheur) 0/3S) ; REMOVE (exheur) PROP) (0 (exheur)PROP) (0 (exheur)PROP) ;
Prepare for re-tokenizer (doesn't fuse if ambiguous)
SELECT (PROP) (-1 PROP) ; # SOS Ásia

Hopes for the future

- Evaluation of **disambiguated** PoS/morphology, *in context* and on *running text*.
- Evaluation of **syntactic** tagging in conjunction with PoS/morphology (would solve ADJ/N, ADV/CONJ, focus particle, and a number of other problems in inter-system comparison)
  - though probably creating a load of new ones ….