ATTEMPT TO FIND A COHERENT SET OF DEFINITIONS
Work done during:
WORD SEGMENTATION session (20 Jan 2006)
and THEMATIC DOMAIN GROUP 2: MORPHO-SYNTAX session (21 Jan 2006)
Jeju (Korea)

Gil FRANCOPOULO (INRIA: LIRICS project)
with the fruitful comments coming from Kiyong Lee (Korea), Maosong Sun (China), Chu-ren
Huang (Taïwan) and Kara Warbuton (Canada)

word
linguistic unit composed of at least a /part of speech/ and a lemma

NOTE: A word is either a single or a compound word. The description can be more complete
with more morphological information and/or syntactic and semantic information

single word
word that does not contain any other word

compound word
word that contains other words

Note: A compound word is both a word and a MWE

MWE
group of words that either:
- has properties that are not predictable from the properties of the individual words or their
normal mode of operation
- are governed by a specific pattern.

Note: A MWE can be a compound word, a fragment of a sentence or a sentence. The group of
words making up an MWE can be continuous or discontinuous. It is not always possible to
mark a MWE with a /part of speech/ information.

Example: A group of words that has properties not predictable from the properties of the
individual words is for instance: "to be over the moon" that means something different from
what it appears to mean. Groups of words governed by a specific pattern are for instance:
"apple pie", "pear pie" with respect to the pattern "<fruit> pie".

morpheme
smallest meaningful sequence of letters, pictograms and numerals
note other proposition from Kiyong: smallest meaningful sequence of linguistics symbols

lexicon
resource comprising words, MWE and affixes

lemma
lemmatised form
conventional form chosen to represent words or MWE

Note: same as in LMF rev-7

form

sequence of morphemes and affixes

affix

same definition as LMF rev-7 i.e. morpheme added to …