questions for POS tagging:
first, we have to decide what tagset to use, then we need to have tests to determine the correct tag for a word in the text.

The following questions are based on the tagset we sent to you a few days ago.

1. Classifiers, Measure words, and nouns

   some systems distinguish classifiers (e.g. 个) from measure words (e.g. 公里). Do we really need two separate tags?

   also, if we merge classifiers and measure words, are the following classifiers or nouns?
   年 in 17 年, 天 in 17 天(note: 月 follows 个 in 17 个 月, and 年 normally cannot stand alone as an NP)

   分 in 得 了 ２９０・１９ 分

2. Localizers (LC) or nouns?

   A Localizer normally follows a noun in the pattern "PREP + NP + LC",

   but some nouns and LC's become kind of fixed combinations, e.g. 海边 窗外 面前
   should they be segmented? if not, should they be tagged as LC or Noun or other?

   also, are 里面 上面 周围 nouns or LCs?

   some LCs can follow sentences, do we need to further classify LCs?

3. NN or NV?
   are noun verbals a separate class of nouns or are they the same as NN except that the words can also be tagged as verbs in some context?

   if they are different tags, do they have different distributions?

   some examples: NN or NV?
   合作 in 各 国 的 经济 贸易 合作
   稳定 和 发展 in 中国 的 稳定 和 发展
   稳定 in 保持 社会 稳定
   促销 in 加大 促销 力度
   忽视 in 不 能 有 任何 忽视

贫困 in 是 贫困 的 症状

4. NN or NR:
   proper nouns as noun modifiers, should they be tagged as NRs or NNs?

   德国 in 德国 军队
   加拿大 in 加拿大 工商界

   can a proper noun be modified by number + classifier?

   全国人大 in 七 届 全国人大 is NN or NR? Note it can be modified by number + classifier

5. determiner or quantifier or predeterminer ?

   do we need a predeterminer for words like 所有 in 所有 这些 人?

   do we need to distinguish determiner from quantifier?
   (some systems use the term quantifier, but they look very much like determiners)

   some examples:
   前 in 前 ６ 轮 => 前 is det?
   很多 in 很多 人 => 很多 is det?
   其 宗旨 => 其 is det or pron?
   此 次 => 此 is det or pron?
   多 名 选手 => 多 is det or number?
   近 ３４ 亿 元 => 近 is adv?

6. verbs or nouns?
   (1) as noun modifiers in "w + Noun":
      管理 in 进行 管理 手段 的 改革
      拚搏 in 有 拚搏 精神 => verb or noun or adj?
      流入 in 资本 大量 流入 情况 下
      精简 in 结合 改革 、 精简 机构 => verb or adj or noun?

   (2) as the object of a verb in "Verb + w":
      改革 in 进行 改革
      应用 in 开始 应用
      发展 in 推进 两 国 关系 的 进一步 发展
         => 进一步 is adv or adj ?

   (3) as head of NP(if it is a noun) or S(if it is a verb)
      批准 in 经 海关 批准
      出口 in 第 一 监狱 向 美 出口 案
      配置 in 实现 棉花 资源 合理 配置 的 流通 体制
      会见 in 会见 时
      介绍 in 据 介绍
      独立 in 一 个 独立 不久 的 国家
      流入 in 国际 资本 流入 的 地区

(4) as head of NP or S in "N de1 w":
    统一 in 爱尔兰 全 岛 的 统一
    欢迎 in 对   中国 贵宾 的 热烈 欢迎
    繁荣 in 德宏 的 繁荣 来自 开放
    危险 in 暴力行动 的 危险

7. Adj or NN or affix or others when modifying NN?

(1) unlike most nouns, they can not be the head of a NP.
unlike most adj, they can not be predicate, and be modified by adv.

部分 in 部分 成果
国际 in 国际 资本 流入 的 地区
重点 in 重点 地区
长期 in 长期 走势 问题
热点 in 热点 问题
大型 in 大型 文艺 表演
宏观 in 宏观 调控
同类 in 同类 国家 公务员 水平 => 同类 is A?
优质 in 优质 服务
人工 in 人工 除草
程控 in 程控 交换机
合资 in 合资 企业

(2) a word has two tags: adj/noun, how to decide which to use in a text, e.g.

传统 in 传统 的 工作 方式
现代化 in 现代化 的 铁路

8. modal or adverb?
将
可能
可 in 可 与 世界 上 著名 的 ...

9. temporal noun or adverb?
今后
迄今 in 迄今 规模 最 大 的
早日
目前
当前
今天

10. Conj or adv
既 in 民兵   既 是 民 又 是 兵
虽 in 关系 虽 有 较 大 发展

11. Noun or adv

直线 in 直线 上升
历史 in 历史 地 看 问题

12. Adj or adv
主要 in 主要 包括
成功 in 研制 成功 的
同步 发展 => 同步 is RB
多 in 更 多 地 发展
高兴 in 高兴 地 说

13. Conj or prep:
easy cases:
与 in 加强 与 世界 各国 的 合作
和 in 伏明霞 和 池彬 分别 ..

ambiguious cases:
与 in 俄 与 拉脱维亚 和 爱沙尼亚 达成 的
与 in 以 加强 日本 与 这些 国家 的 经济 联系

Conj or prep?
经 in 经 北京 海关 批准
由于 in 由于 残疾 种类 的 不同

14. Modal or Verb
需要 in 需要 借鉴 中国 的 经验
愿意 in 愿意 发展 友好合作
需要 in 需要 人们 有所 行动

15. prep or verb

作为 in 中国 作为 联合国 常任 理事国 在 国际 事务 中
发挥 著 特殊 的 积极作用

16. adv or verb?
联合 in 联合 举行 的 早餐会

17. how to tag foreign words?
ＥＶＲＥＣ in 泰国 ＥＶＲＥＣ 国际 有限公司

99: others
兑 in １ 美元 兑 ８．７ 元 人民币
等 in 成套设备 等 资本性 商品
所谓 in 所谓 “ 中国 向 美 出口 案"
是 and 的 in 是 严格 执行 的 :
来 in 来 强化 学习
而 in 为 ... 而 努力 => 而 is Conj?
以 in 以 使 两 国 立法 机构 .. => 以 is Conj?
至 in 五百 至 六百

所 in 所 不 能 接受 的
有所 in 有所 行动 => should 有所 be a word and be tagged as verb?
之 in 运动会 之 事 => 之 is MK1?
其中 => RB
来 in 近 几 年 来 => 来 is LC? 近 is det or adv?
这样 每 平方米 的 => 这样 is RB?