1. n (체언)
   1. nc (보통 명사) 🡪 (NNG)
      1. ncp (서술성 명사) 🡪 (NNG)
         1. ncpa (동작성 명사) 🡪 (NNG)
         2. ncps (상태성 명사) 🡪 (NNG)
      2. ncn (비서술성 명사) 🡪 (NNG)
   2. nq (고유 명사) 🡪 (NNP)
   3. nb (의존 명사) 🡪 (NNB)
      1. nbu (단위성 의존 명사) 🡪 (NNB)
      2. nbn (비단위성 의존 명사) 🡪 (NNB)
   4. np (대명사) 🡪 (NP)
      1. npp (인칭 대명사) 🡪 (NP)
      2. npd (지시 대명사) 🡪 (NP)
   5. nn (수사) 🡪 (NR)\* 수사는 NR 또는 SN으로 매칭 될 수 있음.
      1. nnc (양수사) 🡪 (NR)
      2. nno (서수사) 🡪 (NR)
2. p (용언)
   1. pv (동사) 🡪 (VV)
      1. pvg (일반 동사) 🡪 (VV)
      2. pvd (지시 동사) 🡪 (VA)
   2. pa (형용사) 🡪 (VA)
      1. paa (성상 형용사) 🡪 (VCN)
      2. pad (지시 형용사) 🡪 (VA)
   3. px (보조용언) 🡪 (VX)
3. m (수식언)
   1. mm (관형사) 🡪 (MM)
      1. mma (성상 관형사) 🡪 (MM)
      2. mmd (지시 관형사) 🡪 (MM)
   2. ma (부사) 🡪 (MA)
      1. mag (일반 부사) 🡪 (MAG)
      2. mad (지시 부사) 🡪 (MAG)
      3. maj (접속 부사) 🡪 (MAJ)
4. i (독립언)
   1. ii (감탄사) 🡪 (IC)
5. j (관계언)
   1. jc (격조사) 🡪 (JK)
      1. jcs (주격 조사) 🡪 (JKS)
      2. jco (목적격 조사) 🡪 (JKO)
      3. jcc (보격 조사) 🡪 (JKC)
      4. jcm (관형격 조사) 🡪 (JKG)
      5. jcv (호격 조사) 🡪 (JKV)
      6. jca (부사격 조사) 🡪 (JKB)
      7. jcj (접속격 조사) 🡪 (JC)
      8. jct (공동격 조사) 🡪 (JC)
      9. jcr (인용격 조사) 🡪 (JKQ)
   2. jp (서술격 조사) 🡪 (VCP)
   3. jx (보조사) 🡪 (JX)
      1. jxc 🡪 (JX)
      2. jxf (안씀)
6. e (어미) 🡪 (E)
   1. ef (종결어미) 🡪 (EF)
   2. ep (선어말어미) 🡪 (EP)
   3. ec (연결어미) 🡪 (EC)
      1. ecc (대등적 연결어미) 🡪 (EC)
      2. ecs (종속적 연결어미) 🡪 (EC)
      3. ecx (보조적 연결어미) 🡪 (EC)
   4. et (전성어미)
      1. etn (명사적 어미) 🡪 (ETN)
      2. etm (관형사적 어미) 🡪 (ETM)
7. x (접사)
   1. xp (접두사) 🡪 (XP or XPN)
   2. xs (접미사) 🡪 (XS)
      1. xsn (명사 파생 접미사) 🡪 (XSN)
      2. xsv (동사 파생 접미사) 🡪 (XSV)
      3. xsm (형용사 파생 접미사) 🡪 (XSA)
      4. xsa (부사 파생 접미사) 🡪 (\* + xsa 🡪 MAG)
8. s (기호)
   1. sp (쉼표, 가운뎃점, 콜론, 빗금) 🡪 (SP)
   2. sf (마침표, 물음표 느낌표) 🡪 (SF)
   3. sl (여는 따옴표 및 묶음표) 🡪 (SS)
   4. sr (닫는 따옴표 및 묶음표) 🡪 (SS)
   5. sd (이음표) 🡪 (SS)
   6. se (줄임표) 🡪 (SE)
   7. su (단위 기호) 🡪 (SW)
   8. sy (기타 기호)
9. f (외국어) 🡪 (SL)

문제들

* pvd 문제
  + 이러한: 이러/pvd + ㄴ/etm (kaist), 이러하/VA + ㄴ/ETM (sejong)
  + 그래도: 그리하/pvd + 어도/ecs (kaist), 그래도/MAJ (sejong)
  + 그러면: 그러/pvd + 면/ecs (kaist), 그러면/MAJ (sejong)
* jxf 문제
  + 생각해요: 생각/ncpa + 하/xsv + 어/ef + 요/jxf (kaist), 생각/NNG + 하/XSV + 여요/EF (sejong)
  + 않아요: 않/px + 아/ef + 요/jxf (kaist), 않/VX + 아요/EF (sejong)
  + ~이요,: 능력/ncn + 이/jp + 요/jxf (kaist), 근간/NNG + 이/VCP + 요/EC (sejong)
  + e + jxf 🡪 EF or jxf 🡪 EF or jxf 🡪 EC
* xsa 문제
  + 완전히: 완전/ncn + 히/xsa (kaist), 완전히/MAG (sejong)
  + 분명히: 분명/ncn + 히/xsa (kaist), 분명히/MAG (sejong)
  + Kaist Corpus에서는 ‘분명히’가 mag로 tagging되어 있는 경우가 아주 가끔 있음 (kaistcorpus\_written\_pos\_or\_generality\_explanation95\_mh2-0441.txt)
  + \* + xsa 🡪 MAG
* XPN (체언 접두사) 🡪 XP
  + 무책임: 무/xp + 책임/ncps (kaist), 무/XPN + 책임/NNG (sejong)
  + 불가능: 불/xp + 가능/ncps (kaist), 불/XPN + 가능/NNG (sejong)
  + 문제는 xp가 XPN이 아니라 그냥 XP로 매핑 될 때가 대부분이라는 것이다.
  + Test set의 태그를 모두 XP로 통일할 것인가? 우리의 output을 자세히 만들 것인가?
* nno, nnc 들 중 한글이 아니라 숫자인 것은 SN으로 매핑됨. (다:다 매핑)
* 한자는 f에 속하나 SH로 매핑된다. (1:다 매핑)

mad는 대체로 MAG로 매핑되나 다음과 같은 경우에는 문제가 있다.

왠지