May 2001

Part of Speech Tagging Guidelines for Penn Korean Treebank

Chung-hye Han
University of Pennsylvania

Na-Rae Han
University of Pennsylvania, nrh@ling.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/ircs_reports

Han, Chung-hye and Han, Na-Rae, "Part of Speech Tagging Guidelines for Penn Korean Treebank" (2001). IRCS Technical Reports Series. 24.
http://repository.upenn.edu/ircs_reports/24


This paper is posted at ScholarlyCommons. http://repository.upenn.edu/ircs_reports/24
For more information, please contact libraryrepository@pobox.upenn.edu.
Part of Speech Tagging Guidelines for Penn Korean Treebank

Abstract
This document describes the Part-of-Speech (POS) tagging guidelines for the Penn Korean Treebank Project. The corpus used for this project consists of around 54,000 words and 5,000 sentences. This document starts with a summary of the tagset used in the Penn Korean Treebank, followed by a more detailed discussion of each tag with examples. Then pairs of tags that are easily confused with each other are discussed and guidelines on how to distinguish one from the other for a given base forms and inflections are presented. The document concludes with a list of specific problematic examples with guidelines on how to handle such cases.

Comments
Part of Speech Tagging Guidelines for Penn Korean Treebank

Chung-hye Han, Na-Rae Han

May, 2001
Contents

1 Introduction 1

2 Summary of Tagset 2
2.1 Content Tags ................................................. 2
2.2 Function Tags ................................................ 3
2.3 Symbols .................................................. 4

3 List of Part of Speech with Corresponding Tag 4
3.1 Proper Noun (NPR) ............................................ 4
3.2 Common Noun (NNC) ......................................... 5
3.3 Dependent Noun (NNX) ....................................... 7
3.4 Pronoun (NPN) ............................................... 8
3.5 Number (NNU) ............................................... 9
3.6 Foreign Word (NFW) ........................................ 10
3.7 Verb (VV) .................................................. 10
3.8 Adjective (VJ) .............................................. 10
3.9 Auxiliary Predicate (VX) .................................... 11
3.10 Verbal Adverb, Clausal Adverb (ADV) ..................... 12
3.11 Conjunctive Adverb (ADC) ................................ 13
3.12 Adnominal (DAN) .......................................... 13
3.13 Interjection (IJ) .......................................... 14
3.14 List (LST) .................................................. 14
3.15 Case Postposition (PCA) ................................... 14
3.16 Adverbal Postposition (PAD) ................................ 15
3.17 Conjunctive Postposition (PCJ) ............................ 16
3.18 Auxiliary Postposition (PAU) ................................ 16
3.19 Copula (CO) .................................................. 17
3.20 Final Ending (EFN) ........................................ 17
3.21 Coordinate, Subordinate, Adverbial, Complementizer Ending (ECS) ........................................... 18
3.22 Auxiliary Ending (EAU) .................................... 19
3.23 Adnominal Ending (EAN) ................................... 19
3.24 Nominal Ending (ENM) .................................... 20
3.25 Pre-Final Ending (EPF) .................................... 20
3.26 Suffix (XSF) ............................................... 20
3.27 Verbalization Suffix (XSV) ................................ 22
3.28 Adjectivization Suffix (XSJ) ................................ 23
3.29 Prefix (XPF) .................................................. 23
3.30 Symbols (SCM, SFN, SLQ, SRQ, SSY) ....................... 24

4 Confusing Tags 25
4.1 NNC or ADV .................................................. 25
4.2 NPN or ADV .................................................. 25
4.3 NPN or DAN ......................................................... 26
4.4 NNX or XSF ......................................................... 26
4.5 VJ or DAN ......................................................... 27
4.6 PAU or PAD ......................................................... 27
4.7 PCJ or PAD ......................................................... 27
4.8 PCJ or PAU ......................................................... 28
4.9 XSF or NNC ......................................................... 28
4.10 ADV or DAN ....................................................... 28
4.11 XPF or DAN ....................................................... 29
4.12 ECS or EAU ....................................................... 29
4.13 ENM or EAU ....................................................... 30
4.14 EFN or NNX ....................................................... 30

5 Confusing Examples ............................................. 30
1 Introduction

This document contains the guidelines for annotating Korean texts by part-of-speech, which forms part of the Penn Korean Treebank project. The corpus used for this project consists of texts from military language training manuals. These texts contain information about various aspects of the military, such as troop movement, intelligence gathering, and equipment supplies, among others. Most of the examples and the linguistic issues addressed in this document come from the corpus.

Korean is an agglutinative language with a very productive inflectional system. Inflections include postpositions, suffixes and prefixes on nouns, and tense morphemes, honorifics and other endings on verbs and adjectives. To reflect these characteristics of Korean morphology, the Penn Korean Treebank uses two major types of part-of-speech tags: content tags and function tags. For each word phrase, where a word phrase refers to a fully inflected lexical form, the base form (stem) is given a content tag, and its inflections are given a function tag. Word phrases are separated by a space, and within a word phrase, the base form and inflections are separated by a plus sign (+). The entire tagset used in the Penn Korean Treebank is listed in §2. The main criterion for tagging is syntactic distribution: i.e., a word may receive different tags depending on the syntactic context in which it occurs. For example, ‘있다’ is tagged as a noun if it modifies another noun, and is tagged as an adverb if it modifies a predicate.

This document is organized as follows. §2 presents a summary of the tagset used in the Penn Korean Treebank. In §3, each tag is discussed in more detail with examples. §4 lists pairs of tags that are easily confused with each other and gives guidelines on how to distinguish one from the other for a given base forms and inflections. §5 contains a list of specific problematic examples with guidelines on how to handle such cases.

*We are extremely grateful to Martha Palmer for her continued support and encouragement. We also thank Aravind Joshi, Tony Kroch and Fei Xia for valuable discussions on many occasions. Special thanks are due to Owen Rambow, Nari Kim, and Juntae Yoon for discussions in the initial stage of the project. We also acknowledge Eon-Suk Ko and Mark Dras for comments on the document. The work reported in this document was supported by contract DAAD 17-99-C-0008 awarded by the Army Research Lab to CoGenTex, Inc., with the University of Pennsylvania as a subcontractor, NSF Grant - VerbNet, IIS 98-00658, and DARPA Tides Grant N66001-00-1-8915.
## 2 Summary of Tagset

### 2.1 Content Tags

<table>
<thead>
<tr>
<th>Category</th>
<th>Tag Description</th>
<th>Tag Label</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun</td>
<td>proper noun</td>
<td>NPR</td>
<td>한국 (Korea), 올린턴 (Clinton)</td>
</tr>
<tr>
<td></td>
<td>common noun</td>
<td>NNC</td>
<td>학교 (school), 컴퓨터 (computer)</td>
</tr>
<tr>
<td></td>
<td>dependent noun</td>
<td>NNX</td>
<td>것이 (thing), 등 (etc), 년 (year), 달러 (dollar), 적 (situation)</td>
</tr>
<tr>
<td></td>
<td>personal pronoun, demonstrative pronoun</td>
<td>NPN</td>
<td>그 (he), 이 (this), 무엇 (what)</td>
</tr>
<tr>
<td></td>
<td>ordinal, cardinal, numeral</td>
<td>NNU</td>
<td>먼저 (one), 첫째 (first), 1, 세 (three)</td>
</tr>
<tr>
<td></td>
<td>words written in foreign characters</td>
<td>NFW</td>
<td>Clinton, computer</td>
</tr>
<tr>
<td>predicate</td>
<td>verb</td>
<td>VV</td>
<td>가 (go), 읽 (eat)</td>
</tr>
<tr>
<td></td>
<td>adjective</td>
<td>VJ</td>
<td>pretty (pretty), 다른 (different)</td>
</tr>
<tr>
<td></td>
<td>auxiliary predicate</td>
<td>VX</td>
<td>있 (present progressive), must (must)</td>
</tr>
<tr>
<td>adverb</td>
<td>constituent adverb, clausal adverb</td>
<td>ADV</td>
<td>매우 (very), 조용히 (quietly), 절발 (please), 만일 (if)</td>
</tr>
<tr>
<td></td>
<td>conjunctive adverb</td>
<td>ADC</td>
<td>그 리고 (and), 그 럼 램 (but, however), 및 (and), 또는 (or)</td>
</tr>
<tr>
<td>adnominal</td>
<td>configurative, demonstrative</td>
<td>DAN</td>
<td>새 (new), 연 (old), 그 (that)</td>
</tr>
<tr>
<td>interjection</td>
<td>exclamation</td>
<td>IJ</td>
<td>아 (ah)</td>
</tr>
<tr>
<td>list</td>
<td>list marker</td>
<td>LST</td>
<td>a, (b), 1, 2.3.1, 아, 나.</td>
</tr>
</tbody>
</table>
## 2.2 Function Tags

<table>
<thead>
<tr>
<th>Category</th>
<th>Tag Description</th>
<th>Tag Label</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>postposition</td>
<td>case</td>
<td>PCA</td>
<td>가/이 (nominative), 을/를 (accusative), 의 (possessive), 야 (vocative)</td>
</tr>
<tr>
<td>adverbial</td>
<td>PAD</td>
<td></td>
<td>예시 (from), 토 (to)</td>
</tr>
<tr>
<td>conjunctive</td>
<td>PCJ</td>
<td></td>
<td>약/과, 약고 (and)</td>
</tr>
<tr>
<td>auxiliary</td>
<td>PAU</td>
<td></td>
<td>양 (only), 토 (also), 는 (topic), 마지 (even)</td>
</tr>
<tr>
<td>copula</td>
<td></td>
<td>CO</td>
<td>이 (be)</td>
</tr>
<tr>
<td>ending</td>
<td>final</td>
<td>EFN</td>
<td>는/는/는 (declarative), 니, 는/는, 는/는 (interrogative), 이/네/서 (imperative), 꾸/꾸 (propositive), 구/구 (exclamatory)</td>
</tr>
<tr>
<td>coordinate, subordinate, adverbial, complementizer</td>
<td>ECS</td>
<td></td>
<td>과 (and), 둘 (because), 게 (attaches to adjectives to derive adverbs), 탁고 (that), 탁고 (that)</td>
</tr>
<tr>
<td>auxiliary</td>
<td>EAU</td>
<td></td>
<td>약, 게, 지, 고 (on verbs or adjectives that immediately precede auxiliary predicates)</td>
</tr>
<tr>
<td>adnominal</td>
<td>EAN</td>
<td></td>
<td>는/는 (on main verbs or adjectives in relative clauses or complement clauses of a complex NP)</td>
</tr>
<tr>
<td>nominal</td>
<td>ENM</td>
<td></td>
<td>기, 음 (on nominalized verb)</td>
</tr>
<tr>
<td>pre-final ending (tense, honorific)</td>
<td>EPF</td>
<td></td>
<td>껍 (past), 시 (honorific), 꼬 (future)</td>
</tr>
<tr>
<td>affix</td>
<td>suffix</td>
<td>XSF</td>
<td>남, 들, 적</td>
</tr>
<tr>
<td>prefix</td>
<td>XPF</td>
<td></td>
<td>제, 각, 해</td>
</tr>
<tr>
<td>verbalization suffix</td>
<td>XSV</td>
<td></td>
<td>마다, 되, 시켜</td>
</tr>
<tr>
<td>adjectivization suffix</td>
<td>XSJ</td>
<td></td>
<td>소립, 달, 앱</td>
</tr>
</tbody>
</table>
2.3 Symbols

<table>
<thead>
<tr>
<th>Category</th>
<th>Tag Description</th>
<th>Tag Label</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>comma</td>
<td>sentence ending</td>
<td>SCM</td>
<td>,</td>
</tr>
<tr>
<td>termination</td>
<td>markers</td>
<td>SFN</td>
<td>. ? !</td>
</tr>
<tr>
<td>left quotation</td>
<td>SLQ</td>
<td></td>
<td>’ ( “ {</td>
</tr>
<tr>
<td>right quotation</td>
<td>SRQ</td>
<td></td>
<td>’ ) ” }</td>
</tr>
<tr>
<td>symbols</td>
<td>others</td>
<td>SSY</td>
<td>... ; : -</td>
</tr>
</tbody>
</table>

3 List of Part of Speech with Corresponding Tag

3.1 Proper Noun (NPR)

- Person names, company names, and country names are tagged as a proper noun (NPR).
- A way of distinguishing proper nouns (NPR) from other types of nouns is to use the plural marker ‘들’. Proper nouns cannot take the plural marker ‘들’, whereas other types of nouns can, in general.

(1) Proper nouns (NPR)
   a. *한국들
      Korea-Plural
   b. *홍길동들
      Hongkiltong-Plural

(2) Common nouns (NNC), personal pronouns (NPN)
   a. 전통들
      building-Plural
   b. 그들
      he-Plural

- Proper names with a foreign origin but written in Hangul are tagged as a proper noun (NPR), not as a foreign word (NFW).

엘린도/NPR

- If a code name is composed of a noun and a number, tag the noun as NPR, and the number as NNU.

정산/NPR 16/NNU
3.2 Common Noun (NNC)

- Common nouns constitute the largest subclass of nouns. The examples below illustrates the diversity of this catch-all subclass of nouns.

집, 작전기, 오리, 사람, 동물, 장관, 학약버치, 연어약, 낙서, 강산

- The stem of verbs that are derived with verbalization suffix, and the stem of adjectives that are derived with adjectivization suffix are tagged as common nouns (NNC).

수명/NNC막/XSV막
목막/NNC막/XSV막
노래/NNC시막/XSV막

여성/NNC소리를XSJ막
여성/NNC말/XSJ막
필요/NNC막/XSJ막

- Common nouns that have a foreign origin but are written in Hangul are tagged as common nouns (NNC), not as foreign words (NFW).

예트-106/NNC
결투터/NNC

- Some common nouns include hyphen and a number. These are simply tagged as common nouns (NNC), and are not separated into symbol (SSY) and number (NNU).

예트-106/NNC
승리-80/NNC

- Some lexical items with a temporal meaning are tagged as a common noun (NNC).

어제, 오늘, 내일, 오랫동안, 지금, 동시

- ‘어젯밤, 오늘밤, 내일밤’ are treated as one word and tagged as NNC. But if there is a space between ‘어제, 오늘, 내일’ and ‘밤’, then treat them as separate words and tag accordingly.

어제 밤: 어제/NNC 밤/NNC

- ‘아투, 여룰, 색물, ...’ are tagged as NNC.
• '의우, 그우, 이전, 그전' and '의배, 그배, 저배, 이변, 지변' are each treated as one word and are tagged as NNC.

In the corpus, there may be a space in between '의, 그, 저' and '의배, 변'. In this case, '의, 그, 저' should be tagged as DAN and '의배, 변' should be tagged as NNC and '변' should be tagged as NNX.

의 배: 의/DAN 변/NNC
저 변: 지/DAN 변/NNX

• '그날, 이날, 오늘, 날' are tagged as NNC as a whole.

• '맨' is tagged as NNC, not NNX, because it can be used as an independent noun.

• '전' (before), '후' (after), '상' (above), '하' (below), '중' (midst), '내' (inside), '외' (outside) are all tagged as NNC.

• '이상, 이하' are tagged as NNC.

• '사이, 전적' are tagged as NNC; they are not dependent nouns.

 사이/NNC+가/PCA 벌어졌따
우리 사이/NNC

• '가짜, 참말, 거짓말, 정말, 진짜' are tagged as NNC.

But note that '정말', '진짜' can be tagged as ADV depending on the syntactic context. See §4.1.

• '형' (form) should be separated out and be tagged as a common noun (NNC).

HAK-24/NFW 형/NNC

But '형' (form) in '모형, 비형기' should not be separated from '모'; the whole word '모형' should be tagged as NNC.

Also '형' in '구형, 신형' should not be separated out; '구형, 신형' are words in their entirety and therefore should be tagged as NNC.

• '단락' is tagged as NNC.

109/NNU 연대/NNC 단락/NNC

• '가락' is tagged as NNC.

가락/NNC 색항/NNC

• '편' (different team), '먼' (surface, side), '가' (edge) are tagged as NNC.
### 3.3 Dependent Noun (NNX)

- Dependent nouns are those which cannot occur without the help of a preceding clause, demonstrative, or noun. Some examples include:

  무ropa, 즐, 적, 급, 곡, 줄, 봉, 동, 수, 안, 지, 무문, 테로

Dependent nouns include unit words and classifiers which are preceded by number nouns.

미식, 말락, 개, 볼, 기, 단, 줄, 주, 주일, 권, 명, 마력, 시간, 폰토미식, 센릭, 가직, 균택

But note that for ‘일, 일, 년’, there are cases in which they should be tagged as NNC.

년 /NNC+간/XSF
일 /NNC 2 외
일 /NNC 3 외

Foreign unit words can also be used as an independent noun, when they should be given the NNC tag:

말락/NNC 가직/NNC
미식/NNC 제도/NNC

- ‘차, 외’ followed by a number are tagged as NNX.

제2 차: 제/XPF+2/NNU 차/NNX
제2 외: 제/XPF+2/NNU 외/NNX

- ‘시’ (time) is treated as a dependent noun (NNX), not as a suffix (XSF). In principle, there has to be a space before ‘시’ after the modifying noun or clause, an orthographical convention that is seldom observed. Insert a space if there is none.

전투/NNC 시/NNX
평 외/NNC 시/NNX

But note that ‘시’ in ‘전 시’ is not separated out: ‘전 시’ is a single common noun.

- ‘것, 봉’ in ‘것마’ and ‘봉마’ are NNX.
• When the dependent noun ‘ 걷, 택’ takes other bound morphemes, phonological contraction tends to take place. In this case, ‘ 걷, 택’ and the morphemes should be separated out and tagged accordingly:

간 가: 걷/NNX+의/CO+ㄴ/EFN

간: 걷/NNX+의/PCA

었을 텐데: 걷/NNX+의/CO+ㄴ/텍/EFN

• ‘한번’ should be tagged as NNC because it is almost frozen as a single word. But ‘두 번, 세 번, ...’ should be separated out and ‘변’ should be tagged as NNX. In case when there is no space between the number word and ‘변’, insert a space in between.

한번/NNC
두/NNU 번/NNX
세/NNU 번/NNX

• ‘달’ (month) in ‘한 달, 두 달, 세 달’ and ‘주일’ (week) should be tagged as NNX.

한/NNU 달/NNX
두/NNU 달/NNX
세/NNU 달/NNX
일/NNU 주일/NNX

If there is no space between the number word and ‘달, 주일’, insert a space in between.

• ‘박’ is a dependent noun and should be tagged as NNX.

말 속 드린 밥/NNX+먹/PAD 갈 음/ADV

3.4 Pronoun (NPN)

• Pronouns include personal pronouns, interrogative-indefinites (so-called ‘wh-words’) and demonstrative pronouns.

• ‘의것, 그것, 것것, 것기, 것기, 이것, 그것, 그것, 것것’ are each treated as one word and tagged as pronouns (NPN).

• Some interrogative-indefinites are tagged as pronouns (NPN):

누구, 무엇, 언제, 어디, 얼마나, 약무

• ‘누구나, 무엇이냐, 약무냐’ are tagged as NPN as a whole.
• ‘왜, 얼마나, 얼마나 지, 언제나, 언제든지, 어떻게’, on the other hand, are adverbs and should be tagged as ADV.

• Note that ‘아무’ can be used as an adnominal. In this case, it should be tagged as DAN. See §4.3 for more explanation.

• ‘귀관, 귀악’ are polite second person pronoun forms; they are tagged as NPN.

• ‘자구’ is NPN.

• When ‘어민가/어디인가, 누군가/누군인가’ correspond to ‘somewhere, someone’ in meaning, they are recognized as single words and should be tagged as NPN as a whole:

어민가로 갓다: 어민가/NPN+로/PAD
누군가를 만났다: 누군가/NPN+를/PCA

However, in the following examples, they should be treated as predicative nominals, and be broken into base forms and copula:

작내 집이 어민가?: 어민가/NPN+이/CO+\+가/EFN
저 사람의 누군가 모르겠다: 누군/NPN+이/CO+\+가/EFN

3.5 Number (NNU)

• Number words (NNU) include cardinals (‘착하, 들, 만, 두’), ordinals (‘첫째, 둘째’) as well as numbers (‘1, 2, 3’). NNU is the only category that is based on a semantic feature rather than syntactic distribution: ‘착하’ and ‘만’ both receive the NNU tag since they are numeric in terms of meaning, although syntactically they are quite different in that ‘착하’ functions as a noun and ‘만’ as a prenominal modifier.

• ‘두 역, 서너, 너댓...’ are also NNU.

• ‘몇’ is considered to be numeric in meaning and is therefore tagged as NNU, not DAN.

몇 가지: 몇/NNU 가지/NNX

But ‘여칠’ is one word and is tagged as NNC.
3.6 Foreign Word (NFW)

- Foreign words include any words not written in Hangul. For instance, ‘Clinton’ is tagged as NFW, but ‘클린턴’ is tagged as NPR.
- In cases like ‘R(에르)-116’ and ‘HAK(하크)-24,’ erase the material in parenthesis and tag them as NFW.

\[\begin{align*}
R(에르)-116: & \quad \text{R-116/NFW} \\
Hak(하크)-24: & \quad \text{H-24/NFW}
\end{align*}\]

- Unit words written in Roman characters are tagged as NFW.

\[
\begin{align*}
\text{mm/NFW} \\
\text{km/NFW}
\end{align*}
\]

3.7 Verb (VV)

- Verbs are inflected for tense, honorific, mood, sentence types, and they serve as a predicate. Verbs are either main or auxiliary. The tag VV are only assigned to main verbs.

3.8 Adjective (VJ)

- Adjectives are inflected essentially the same way as verbs, and they function as predicates.
- There are some formal differences between adjectives and verbs. While verbs can take present tense marker ‘는/ㄴ’, adjectives cannot:

\[
(3) \quad \begin{align*}
a. & \quad \text{철수 가 뛰고 있다}. \\
& \quad \text{Cheolswu-Nom run-Pres-Decl} \\
& \quad \text{‘Cheolswu is running.’} \\
b. & \quad * \quad \text{꽃이 예쁘다}. \\
& \quad \text{flower-Nom pretty-Pres-Decl} \\
& \quad \text{‘The flower is pretty.’}
\end{align*}
\]

While verbs can take a progressive auxiliary predicate, adjectives cannot:

\[
(4) \quad \begin{align*}
a. & \quad \text{철수 가 뛰고 있다}. \\
& \quad \text{Cheolswu-Nom run be-Decl} \\
& \quad \text{‘Cheolswu is running.’} \\
b. & \quad * \quad \text{꽃이 예쁘고 있다}. \\
& \quad \text{flower-Nom pretty be-Decl} \\
& \quad \text{‘The flower is being pretty.’}
\end{align*}
\]
• ‘있지, 없지, 계시, 아니지’ are tagged as adjectives (VJ).

• If ‘있지, 계시’ are used as auxiliary predicates, then they should be tagged as VX.

<table>
<thead>
<tr>
<th>prep</th>
<th>V+</th>
<th>EAU</th>
<th>V+</th>
<th>EFN</th>
</tr>
</thead>
<tbody>
<tr>
<td>민</td>
<td>VV</td>
<td>과</td>
<td>EAU</td>
<td>있</td>
</tr>
<tr>
<td>잡수</td>
<td>VV</td>
<td>과</td>
<td>EAU</td>
<td>계서</td>
</tr>
</tbody>
</table>

3.9 Auxiliary Predicate (VX)

• Auxiliary predicates include auxiliary verbs such as ‘농락, 퇴락’ as well as auxiliary adjectives such as ‘있락, 싫다’. They cannot stand alone without a preceding verb or an adjective. Some examples of auxiliary predicates include:

농락, 퇴락, 나락, 오락, 버락락

박구/VV+의/EAU 농/VX+_alloc/EFN
식/VV+고/EAU 낙/VX+식/PCS
먹/VV+계/EAU 채/VX+ALLOC/EPF+铛/EFN
 gameTime/VJ+같/EAU 오/VX+alloc/EPF+铛/EFN
개/VV+의/EAU 버락/VX+铛/EFN

식락, 보락, 아니락락

노태/NNC+의/XSV+고/EAU 식/VX+铛/EFN
노태/NNC+의/XSV+는/가/EFN 식/VX+铛/EFN
嘁/VV+의/EAU 보/VX+alloc/EPF+铛/EFN
嘁/VV+는/가/EFN 보/VX+铛/EFN
낙박/VJ+같/EAU 알/VX+铛/EFN
먹/VV+적/EAU 알/VX+alloc/EPF+铛/EFN

만약락, 둘락락, 볼락락, 젓락락, 짭락락

블 만약락: 보/VV+고/EAU 만약/VX+铛/EFN
우는 둘약락: 올/VV+는/EAN 둘약/VX+铛/EFN
창알 볼약락: 창야/VJ+고/EAN 볼약/VX+铛/EFN
خرج 자른약락: 자른/VJ+고/EAN 자른약/VX+铛/EFN
막칠 짭약락: 짭약/VV+고/EAN 짭약/VX+alloc/EPF+铛/EFN

Sometimes ‘만, 둘, 볼, 젓, 짭’ accompany a postposition and are separated from the verbal stem ‘막’; in these cases they must be tagged as NNX:

블 만도 약락: 보/VV+고/EAU 만/NNX+도/PAU 덤/VV+铛/EFN
• Note that some words are ambiguous between auxiliary predicates and main verbs. If they are used as auxiliary predicates, they are tagged as VX, and if they are used as main verbs, they are tagged as VV.

Verbal Adverb, Clausal Adverb (ADV)

• Adverbs modify either the entire clause (clausal adverbs) or the verb phrase (phrasal adverbs). Clausal adverbs include ‘악마, 제발, 옥시, 과연, 만일, 만약, 비록, 부터, 설마, 타령익, 맘만, 또만, 또’.

Phrasal adverbs describe manner, time, place or degree. They include ‘발리, 천천히, 밵토, 작주, 벌써, 약까, 즉시, 매우’.

Onomatopoeic manner adverbs or phenomimes include ‘소곤 소곤, 탄각탄각, 탄탕탄탕, 악장악장, 번적번적, 성글병글’.

Some other examples include:

오택, 모두, 보통, 통상, 직접, 즉시, 얼막든지, 얼막나, 얼막만금, 어렴게, 어렴게, 그렴게, 저렴게, 역시, 거무로, 방향, 전역, 이짜금, 이짜금씩, 가끔, 가끔씩, 테 부분, 탐락, 탐, 각각

• ‘그리, 그리로, 저리, 저리로’ are tagged as ADV.

• Adverbs derived from an adjective with suffix ‘의’ are tagged as ADV as a whole, without breaking them into the adjective and the suffix.

옆의/ADV
갈의/ADV

• ‘기껏해야’ is a single word and is tagged as ADV.
• ‘도, 도안’ are tagged as ADV, not ADC. But ‘도는’ is tagged as ADC (Conjunctive Adverb). The difference between the two groups is analogous to the one between English ‘also’ and ‘and’: the former can only be used to modify predicates, while the latter can function as a coordination operator that conjoins two words or phrases, as seen below:

철수가 도 시험에 빌어졌담.
철수도안 가약 한담.
점심도는 저녁

• Negation ‘안, 못’, which precede the verb, are tagged as ADV.

안 먹겠다: 안/ADV 먹/VV+겠/EPF+하/EFN
못 먹겠다: 못/ADV 먹/VV+겠/EPF+하/EFN

3.11 Conjunctive Adverb (ADC)

• Conjunctive adverbs include words that coordinate words, phrases or sentences. They may also occur at the beginning of a sentence. Some examples are:

혹은, 또는, 및, 즉, 내지, 그리고, 그러니까, 그러니까가, 그러니까면, 그러니까므로, 왜냐하면

3.12 Adnominal (DAN)

• Adnominals do not conjugate or inflect, and they modify nouns at a pre-nominal position. Some of them have determiner-like properties, hence the tag name Determiner-AdNominal. Examples are:

어, 그, 이, 오, 세, 한, 여력, 약, 안, 무순, 어편, 어편, 저편, 그편, 어느, 모든, 악무, 각, 전, 제, 주, 지난, 별, 막론, 박로

Note that some of these examples have ambiguous usages. For example, 안 is ambiguous between adnominal usage and numeral usage. See §5 for a way to distinguish between the two usages.

Adnominals are different from adjectives in that adnominals can only be a modifier, while adjectives can be used as a modifier as well as a predicate.

(5) Adjective
a. 꽃이 예쁘다.
   flower-Nom pretty-Decl
   ‘The flower is pretty.’

b. 예쁜 꽃
   pretty-Rel flower
   ‘pretty flower’
(6) Adnominal
   a. 새 책
      new book
   b. 이 책
      this book

- Some mono-syllabic Sino-Korean items usually appear as a prefix (XPF), but they are recognized as an adnominal when delimited by a space. These include ‘전’ and ‘주’.

진 군:  진/DAN 군/NNC
진 연체:  진/DAN 연체/NNC
주 전투 전자:  주/DAN 전투/NNC 전자/NNC

3.13 Interjection (IJ)
- ‘예’ should be tagged as IJ. But ‘약拿来, 약拿来’ are viewed as inflected forms of the adjective ‘약拿来’ and therefore should be given VJ tag plus appropriate tags for the inflections.

약拿来/VJ+오/EFN
약拿来/VJ+ㅂ니답/EFN

- ‘글쎄,’ and ‘글쎄요’ are tagged as IJ.

3.14 List (LST)
- Letters and numerals which identify items in a list, together with their surrounding punctuation are tagged as LST.

1./LST
 i./LST
 (a)/LST
 ㄱ./LST
 ㄴ./LST
 2.3.1/LST

3.15 Case Postposition (PCA)
- A case postposition specifies the grammatical function of the noun phrase which it is attached to, e.g., subject, object. Case postpositions include nominative marker (‘ ’), accusative marker (‘을/를’), genitive marker (‘의’), and vocative marker (‘야”).
• The marker ‘/’ on the second noun phrase of double nominative constructions is not treated differently from the ones on subject noun phrase; it is also tagged as case postposition (PCA).

네가 팔에 막بري먹었어:
어/NPN+어/PCA 꺼/NNC+어/PCA 막بري/NNC+어/XSV+였/EPF+amsung

3.16 Adverbial Postposition (PAD)

• Postpositions that correspond to English prepositions in meaning and usage are adverbial postpositions. Some examples include:

예게, 으로, 까지, 부터, 안데, 로부터, 에와가, 에서부터

Note that ‘로부터, 에서부터’ are not broken into two individual postpositions.

• Comparative markers such as ‘처럼, 보다’ are also tagged as PAD.

단 계/NNC+처/NSF+으로/PAD

• ‘따고’ attached to ‘Noun-적’ should be tagged as PAD.

만약/NPR+따고/PAD 는다.

However, the same ‘따고’ receives a different tag, ECS, the “complementizer tag” when it attaches to a predicative noun in a clausal complement. In this case, a copula must be recovered before it, as seen in the following example:

철수는 완작가 물고기라라고 말이는다.
완작/NPR+어/PCA 물고기/NNC+어/CO+따고/ECS

But seemingly similar ‘따는’ in the example below is classified as EAN (Adnominal ending), since the noun along with the ending is the head of the relative clause which modifies the following noun. In this case, a copula must also be recovered before ‘있다’. Refer to the entries for ‘따고’ and ‘있다’ in §5 for more information.

완작따는 물고기
완작/NPR+있다/CO+따는/EAN 물고기/NNC

• ‘안데’ and ‘안데서’ are two different postpositions, and they should each be tagged as PAD. They have different meanings; the former has the meaning of ‘goal’, whereas the latter has the meaning of ‘source’.

수릭병안데: 수릭병/NNC+안데/PAD
수릭병안데서: 수릭병/NNC+안데서/PAD
3.17 Conjunctive Postposition (PCJ)

- Conjunctive postpositions correspond to English ‘and’, or ‘or’. Some examples include:
  약고, 날/이날, 와/와, 탁든지/이탁든지
- Note that these postpositions can also be used as an adverbial postposition. When they are used as an adverbial postposition, they are tagged as PAD. See §4.7 for how to distinguish between conjunctive postpositions and adverbial postpositions.

3.18 Auxiliary Postposition (PAU)

- Auxiliary postpositions occur on nouns in addition to adverbial postpositions. They may also occur on nouns in place of case postpositions. Some examples of auxiliary postpositions include:
  ‘는, 만, 도, 밖에, 마다, 조차, 까지, 부터’.
- ‘(이)란’ which is short for ‘학교 학는 것들’ is tagged as PAU.

Complex postpositions that have undergone phonological contraction should be separated out into individual parts and tagged accordingly:
3.19 Copula (CO)

- The copula marker ‘의’ attaches to nouns and turns them into a predicate.

박보 이다: 박보/NNC+의/CO+가/EFN

In many cases, especially when the preceding noun ends in a vowel, the copula marker may not show up on the surface. A special attention needs to be paid to make sure that the copula is identified in these cases. See the entry for ‘의’ in §5 for further information on how to determine when a copula is needed.

3.20 Final Ending (EFN)

- Final endings occur on predicates and mark sentence types. For instance, they indicate whether a sentence is a declarative, interrogative, imperative, propositive, and so on.

- Note that final endings can also occur in a non-terminal environment: on verbs preceding an auxiliary predicate or on main verbs that occur in complement clauses.

Before an auxiliary predicate:

-는 가/EFN 봤나 막
-는 가/EFN 살락

On verbs in complement clauses:

-는 가/EFN+에/PAD 말았따
-는 지/EFN+을/PCA 못했따
-가지 없/VX+있/EPF+가/EFN 생각안따

- ‘는 지’ by itself is seldom used as a final ending except perhaps in the following cases. However, it patterns with other EFN endings such as ‘는 가, ‘서, ‘려’ in terms of its syntactic distribution; therefore it is tagged as EFN, not ECS.

철수 는 가는 지?  
철수 는 가는 지요?

가는 가 말하탁: 가/VV+는 가/EFN 말하탁
가는 지 말하탁: 가/VV+는 지/EFN 말하탁
가는 가 살락: 가/VV+는 가/EFN 살락
가는 지 살락: 가/VV+는 지/EFN 살락
가는 가를 말하탁: 가/VV+는 가/EFN+을/PCA 말하탁
가는 지를 말하탁: 가/VV+는 지/EFN+을/PCA 말하탁
가는 가가 문제이탁: 가/VV+는 가/EFN+가/PCA 문제이탁
가는 지가 문제이탁: 가/VV+는 지/EFN+가/PCA 문제이탁
### 3.21 Coordinate, Subordinate, Adverbial, Complementizer Ending (ECS)

- Coordinate/subordinate/adverbial/complementizer endings occur on predicates of coordinate adjunct clauses or subordinate adjunct clauses.

- Complementizers occurring on the predicate of a complement clause under verbs such as ‘말약사’, ‘생각할까’, ‘물’, ‘오구할까’ are tagged as ECS. Examples include ‘탁고’ and ‘탁고’. See §3 for further information on the item ‘탁고’.

  그 곳에 전화기가 없잖아 생각할까: 엣/VJ+탁고/ECS
  편지를 쓸ownik 말랬다: 쓰/VV+논탁고/ECS

- Conditional markers ‘탁면’ and ‘탁면’ are tagged as ECS:

  엣탁면: 엣/VJ+탁면/ECS
  가탁면: 가/VV+탁면/ECS

- ‘계’ on adjectives as in ‘억볼 난계’ should be separated out and be tagged as ECS. But ‘계’ in ‘억볼 계’ should not be separated out: the whole lexical item is treated as an adverb (ADV).

  If ‘계’ is followed by an auxiliary predicate, it is tagged as EAU. See §3.22 for an example.

- Ending ‘는께’ must be tagged as ECS.

  However, when ‘는께’ expresses ‘the place which...’, a space should be inserted between ‘는’ and ‘께’, since ‘께’ in this case is a dependent noun and ‘는’ is a part of the relative clause that modifies it.

  질수는 엣/VJ+는께/ECS 영의는 없닭.  (meaning ‘있지만’)
  께이 엣/VJ+는/EAN 계/NNX 남이가 있다.  (meaning ‘있는 곳에’)

- ECS can occur on verbs or adjectives preceding an auxiliary predicate.

  설치/NNC+막/XSV+특목/ECS 되/VX+의/EAU 엣/VX+탁/EFN

  ‘특목’ is ECS since it occurs on main verbs in subordinate clauses.

  일찍 끝내도 특 노력해야.

  Endings ‘억야’, and ‘억도’ on verbs or adjectives that precede modal auxiliary predicates are tagged as ECS.
3.22 Auxiliary Ending (EAU)

- Auxiliary endings occur on verbs or adjectives that are followed by auxiliary predicates. Auxiliary endings include ‘よ,’ ‘げ,’ ‘じ,’ ‘ご,’ their morpho-/phonological variations and nothing else:

- Some auxiliary verbs take a verbal complement with EAN ending. Examples are:

- Some auxiliary verbs take a verbal complement with an ECS ending. Examples are:

3.23 Adnominal Ending (EAN)

- Adnominal endings ‘는, 는, 을, 를, 름’ occur on predicates in a relative clause or complement clause of a complex NP.

- In cases where ‘는’ is a short for ‘는 마고는’, ‘는 름’ should simply be tagged as EAN.

- ‘는’ as in ‘학생이 름는’ is tagged as EAN. Note the presence of the copula ‘의’ between the noun and ‘는’. It is this copula marker that takes the verbal ending, not the noun itself. Often the copula marker is not visible: make sure to put it in.
3.24 Nominal Ending (ENM)

- Nominal endings ‘가, 음’ nominalize the verb/adjective that they attach to.
- A predicate that has been nominalized has the same morpho-syntactic distributional pattern as nouns, and therefore it can take a postposition:

  약관으로써: 약관/NNC+가/XSV+□/ENM+으로써/PAD

3.25 Pre-Final Ending (EPF)

- Pre-Final endings include tense and honorific markers, such as ‘었/었, 것, 싶/은 시’.
- These morphemes can co-occur, as in ‘섰, 엮었, 엮겠, 엮겠, 시겠’. They are not further analyzed, and receive a EPF tag as a whole:

  겪겠습니다: 겪/VV+겠습니다/EPF+닭/EFN
  막겠습니다: 막/VV+했습니다/EPF+닭/EFN

3.26 Suffix (XSF)

- Suffixes mostly attach to nouns. Some examples include:

  닭, 네, 네들, 네들이, 품, 개, 씨, 키, 식, 셋, 덤, 상, 쪖, 식, 맛, 급, 심

  선생님: 선생/NNC+님/XSF
  도의상: 도의/NNC+상/XSF
  이론상: 이론/NNC+상/XSF
  무전기통: 무전기/NNC+통/XSF
  막막한: 막막/NNU+한/XSF
  작가: 작가/NNPN+가/XSF
  10 미터짜리: 10/NNU 미터/NNX+짜리/XSF
  양모의: 양모/NNC+의/XSF
  연대석: 연대/NNC+석/XSF
  직위관적: 직위관/NNC+적/XSF
  시관성: 시관/NNC+성/XSF

- Some suffixes such as ‘들’ can attach to words other than nouns.

  약/NNC+들/XSF
  빨/ADV+들/XSF

Note that ‘들’ can also be used as a dependent noun. See §4.4 for a way to distinguish NNX usage from the XSF usage for ‘들.’
• ‘적’ on nouns is tagged as XSF.

단계/NNC+적/XSF
사위/NNC+적/XSF

But note that ‘비고적’ is always tagged as ADV as a whole, because it can only be used to modify a predicate, whereas other noun+적 items mostly function as a prenominal modifier.

비고적/ADV

• ‘구, 신, 군, 현, 읍, 티, 동’ occur with city, town, place names. These should be tagged as XSF.

• ‘성, 장, 제, 풍, 명, 소, 포, 오, 부, 병, 권’ and ‘께’ are not recognized as suffixes. Nouns occurring with these morphemes should simply be tagged as NNC as a whole.

특무장/NNC
구분대/NNC
가능성/NNC
초전제/NNC
부분품/NNC
통신망/NNC
글씨소/NNC
지위소/NNC
승립-58오/NNC
앞모토/NNC
앞병부/NNC
앞병/NNC
지위권/NNC

• When the annotator is not sure whether a certain item should be tagged as XSF or not, the following strategy applies in general: it is tagged as XSF if it is very productive in a general domain, but it is not recognized as an individual suffix if it can rather be thought of as a part of a domain-specific word. For instance, in the example below, ‘선’ is not tagged as XSF because it is part of domain-specific words ‘삼선, 오선’.

삼선/NNC, 오선/NNC 절초량
양사기조/NNC
만기조/NNC
Verbalization Suffix (XSV)

- Verbalization suffixes attach to nouns to turn them into a verb. They include ‘먹, 떠, 시켜, 벌.’
- Verbs that are derived by a verbalization suffix should be separated into the noun stem and the suffix part and tagged accordingly.

\[
\text{수영/NNC+먹/XSV+먹/EFN} \\
\text{독특/NNC+떠/XSV+떠/EFN} \\
\text{도끼/NNC+끼/XSV+끼/EFN}
\]
- But the following items are viewed as frozen forms, and therefore ‘먹’ should not be separated out:

\[
\text{말먹/VV+먹/EFN} \\
\text{가먹/VV+먹/EFN} \\
\text{못먹/VX+먹/EFN} \\
\text{만먹/VX+먹/EFN} \\
\text{ kötü/VX+먹/EFN}
\]
- ‘먹’ is not a verbalization suffix. If it occurs attached to a noun, the entire item should be tagged as VV. If it is separated from a noun by space, then the noun should be tagged as NNC, and ‘먹’ should be tagged as a verb.

\[
\text{고장먹/고장/VV+먹/EFN} \\
\text{반항먹/반항/VV+먹/EFN} \\
\text{불먹/불/VV+먹/EFN}
\]

\[
\text{고장/NNC 먹/VV+먹/EFN} \\
\text{반항/NNC 먹/VV+먹/EFN} \\
\text{불 먹/NNC 먹/VV+먹/EFN}
\]
- ‘드리’ is not treated as a verbalization suffix either. If it is separated from a noun by space, again the noun should be tagged as NNC, and ‘드리’ as a verb. If there is no space in between, the annotator must put it in. However, some expressions such as ‘말씀드리막, 인사드리막’ are recognized as a word as a whole.

\[
\text{말씀드리막/말씀드리/VV+막/EFN} \\
\text{말씀드리/VV+막/EFN} \\
\text{말씀드리/NNC 드리/VV+막/EFN} \\
\text{선물드리막/선물/NNC 드리/VV+막/EFN}
\]
- ‘받’, however, is recognized as a verbalization suffix when attached to a common noun, because it can be viewed as a grammaticalized means of passivization. Note that it can also function as a main verb, as in the first example below.

\[
\text{보충 받막/보충/VV+막/EFN} \\
\text{보충받막/보충/NNC+받/XSV+막/EFN}
\]
3.28 Adjectivization Suffix (XSJ)

- Adjectivization suffixes attach to nouns to turn them into adjectives. They include ‘스럽, 밝, 악.’

- Adjectives that are derived by an adjectivization suffix should be separated into the stem and the suffix, and tagged accordingly.

\[
\begin{align*}
\text{역성/NNC+스럽/XSJ+막/EFN} \\
\text{역성/NNC+밝/XSJ+막/EFN} \\
\text{투명/NNC+악/XSJ+막/EFN}
\end{align*}
\]

Note that ‘막’ can be tagged as XSV or XSJ depending on how it is used.

3.29 Prefix (XPF)

- Prefixes immediately precede nouns with no intervening space. Some examples include:

\[
\begin{align*}
\text{제, 각, 반, 벌}
\end{align*}
\]

\[
\begin{align*}
\text{제23: 제/XPF+23/NNU} \\
\text{각금: 각/XPF+금/NNC} \\
\text{반공산주의: 반/XPF+공산주의/NNC} \\
\text{벌정규전: 벌/XPF+정규전/NNC}
\end{align*}
\]

Note that ‘제, 각’ can also be used as DAN. See §4.11 for information on how to distinguish the DAN usage from the XPF usage for these items.

- ‘매’ (each) is tagged as XPF.

\[
\begin{align*}
\text{매시관: 매/XPF+시관/NNC} \\
\text{매약기: 매/XPF+약기/NNC}
\end{align*}
\]

But ‘매일’ is tagged as NNC as a whole.

- ‘총’ is not tagged as XPF:

\[
\begin{align*}
\text{총 경영/NNC}
\end{align*}
\]

- In general, when ‘매, 초’ precede a noun, they are tagged as XPF. But some frozen words such as given below are tagged as NNC as a whole.
In general, ‘재’ (again) is tagged as XPF. But some frozen words that are prefixed with ‘재’ are tagged as NNC as a whole.

재 전 개: 재/XPF+전 개/NNC
재 결 합: 재결 합/NNC
재 교 육: 재교육/NNC

When the annotator is not sure whether a certain item should be identified as XPF or not, in general, it should be tagged as XPF if it is very productive, but it should not be tagged as XPF if the entire word can be viewed as a domain-specific word. For instance, in the example below ‘리’ is not tagged as XPF because it is part of the domain-specific word ‘리직원’.

리 직 원/NNC
반 텅 스/NNC
반 항 공/NNC

3.30 Symbols (SCM, SFN, SLQ, SRQ, SSY)

- Parentheses ‘(, )’ conflict with the internal tree-bracket notation used in the Penn Korean Treebank. Therefore they should be rewritten as ‘-LRB-’ and ‘-RRB-’ respectively. They bear the SLQ and SRQ tags.

OZM-3 (오 제 엔-3):
OZM-3/NFW -LRB-/SLQ 오 제 엔-3/NNC -RRB-/SRQ

- Likewise, braces ‘{, }’ are rewritten as ‘-LCB-’ and ‘-RCB-’, and square brackets ‘[, ]’ are rewritten as ‘-LSB-’ and ‘-RSB-’. Left brackets are tagged as SLQ, and right brackets are tagged as SRQ.

- ‘…’ and ‘--’ are each recognized as a single symbol.

.../SSY
---/SSY
4 Confusing Tags

4.1 NNC or ADV

- Some items are ambiguous between ADV and NNC. Some examples include:

  정말, 진짜, 설계, 만약, 만일, 절대, 야까, 매일, 총시, 도랍, 그동안

  The syntactic context in which they occur must be taken into consideration in order to determine which tag should be given.

  If these items modify predicates, they are tagged as ADV:

  정말/ADV 좋악.
  절대/ADV 안원식.
  만약/ADV 간박연
  야까/ADV 꿀약.
  그동안/ADV 일마웠악.

  But if they accompany a postposition and/or modify other nouns, they should be tagged as NNC.

  진짜/NNC+여/PCA 아니약.
  정말/NNC+여/CD+여/EFN
  만약/NNC+여/PAD
  만일/NNC+의/PCA 경우/NNC
  절대/NNC 일도/NNC
  야짜/NNC+의/PCA 약속
  매일/NNC+의/PCA 일과
  그 총시/NNC+예/PAD+는/PAU 제 정신의 아니없악
  그동안/NNC+의/PCA 실직/NNC

  Also note that ‘정말로, 진짜로, 설계로, 절대로’ are not broken into ‘noun’+‘로’ but tagged as ADV as a whole.

4.2 NPN or ADV

- ‘모두’ is ambiguous between NPN and ADV. If it modifies a predicate, it is tagged as ADV. If it is used as a noun, it is tagged as NPN.

  약이들은 모두/ADV 모약약.
  모두/NPN+여/PCA 왔약.
4.3 NPN or DAN

- Some words such as ‘어, 그’ are ambiguous between NPN and DAN. If they occur without any postpositions and modify a noun, they are tagged as DAN. If they carry a postposition, they are tagged as NPN.

4.4 NNX or XSF

- For some items, it may be difficult to determine whether they are NNX or XSF. In general, if they are preceded by a space, they are NNX, and if they are immediately adjacent to the preceding noun, then they are XSF.

- ‘식’ following a predicate with an adnominal ending should be tagged as NNX. ‘식’ in this context means ‘way’, or ‘method.’

‘식’ immediately adjacent to a noun should be tagged as XSF.

- ‘들’ usually attaches to a noun, and marks plural. It can also occur on words other than nouns. In both cases, it is tagged as XSF.

But note that it can also be tagged as NNX in the particular context shown below:

- If ‘간’ is translatable as ‘between’, it is tagged as XSF, but if it is translatable as ‘during, for’, then it is tagged as NNX.
4.5 VJ or DAN

- Note that ‘ vå트 ’ is ambiguous between an adnominal and an adjectival usage. When it is translatable as ‘ other ’, it should be tagged as DAN, but when it is translatable as ‘ different ’, it is viewed as an inflected form of the adjective ‘ vå트 ’ and therefore should be tagged as \( \text{ vå트}/\text{VJ}+\text{EAN} \).

4.6 PAU or PAD

- ‘ 까지, 부터 ’ are ambiguous between adverbial postposition (PAD) and auxiliary postposition (PAU). When they occur with nouns that refer to time or location and are used as adjuncts, they should be tagged as PAD. But when they occur with nouns that are either subject or object of the sentence, then they should be tagged as PAU.

4.7 PCJ or PAD

- ‘ 와/과 ’ is ambiguous between adverbial postposition (PAD) and conjunctive postposition (PCJ). When it is used to conjoin two words or phrases, it is a conjunctive postposition and therefore tagged as PCJ. On the other hand, when it is used to mark an internal argument of certain class of adjectives such as ‘ 갈다, 맨트다 ’, it is an adverbial postposition, and is tagged as PAD.

(7) Conjunctive postposition (PCJ)

a. 너와 나는 \( \text{ 까지, 부터} \).
   ‘You and I are different.’

b. 너와 \( \text{ 까지, 부터} \) 나는 갈다.
   ‘You and I are same.’

(8) Adverbial postposition (PAD)

a. 나는 너와 \( \text{ 까지, 부터} \)。
   ‘I am different from you.’

b. 나는 너와 \( \text{ 까지, 부터} \) 갈다.
   ‘I am same as you.’
4.8 PCJ or PAU

- The postposition ‘나/이너’ is ambiguous between auxiliary and conjunctive postposition. If it conjoins two NPs, then it is PCJ. Otherwise, it is tagged as PAU.

samp| 오렌지 먹었다| 삼광/NNC+나/PCJ 오렌지/NNC+를/PCA
세 번이너 없었다| 세/NNU 변/NNX+이너/PAU

4.9 XSF or NNC

- If ‘과’ or , ‘금’ immediately follows a noun with no intervening space, they are tagged as XSF. But if they are preceded by a space or a prefix, they are tagged as NNC.

영문네: 영문/NNC+과/XSF
좋은 과: 좋은 과/NNC
각금: 각/XPF+과/NNC

좋은 금: 좋아/NNC+금/XSF
금의 막트막: 금/NNC+의/PCA 막트막
각금: 각/XPF+금/NNC

But note that ‘병과’ should be tagged as NNC as a whole. It is a domain-specific word.

병과/NNC

4.10 ADV or DAN

- ‘박또’
  ‘박또’ can be an adverb (ADV) or an adnominal (DAN). When it modifies a verb or an adjective it is an adverb; when it modifies another adnominal or a noun it is DAN.

박또/ADV 맛고로 먹았다.
박또/DAN 적/DAN 사탕을
박또/DAN 케다가

- ‘거의’, ‘DOMContentLoaded’, ‘덕’
  ‘거의’, ‘DOMContentLoaded’ and , ‘덕’ can also be an adverb (ADV) or an adnominal (DAN).

거의/ADV 동났다.
거의/DAN 모든/DAN 공급품의

DOMContentLoaded/ADV 업니다.
DOMContentLoaded/DAN 전방/NNC+에/PAU+를/PAU

덕/ADV 업니다.
덕/DAN 이상/NNC 모든다.
4.11 XPF or DAN

- ‘제’ is tagged as prefix (XPF) if it occurs before a number without any space. If ‘제’ is followed by a space and a number, it is tagged as DAN.

\[ 제23: \text{제/XPF+23/NNU} \]
\[ 제 23: \text{제/DAN 23/NNU} \]

If ‘제’ modifies a common noun and means ‘all’, it should be tagged as DAN.

\[ 제 명제: \text{제/DAN 명제/NNC} \]

- ‘것’, when immediately preceding a noun, should be tagged as XPF.

\[ 것 금: \text{것/XPF+금/NNC} \]

On the other hand, if ‘것’ is followed by a space and a noun, it should be tagged as DAN:

\[ 것 명것의: \text{것/DAN 명것/NNC+의/PCA} \]

But ‘것 중’ is treated as a single word: the entire lexical item is tagged as NNC.

4.12 ECS or EAU

- Some endings are ambiguous between ECS and EAU. For instance, if ‘와’ occurs on a predicate followed by an auxiliary predicate, it is tagged as EAU. If it occurs on a predicate of a subordinate clause, it is tagged as ECS. One way of distinguishing between the two usages is to replace ‘와’ with ‘와 서’: if it is grammatical and preserves the overall meaning, then ‘와’ has a similar function as ‘와 서’, and, therefore, is an ECS.

\[ 먹 와 보았습니다: \text{먹/VV+와/EAU 보/VX+없/EPF+먹/EFN} \]
\[ 긁어 놓았습니다: \text{금/VV+을/ECS 놓/VV+없/EPF+먹/EFN} \quad (= \text{금 약서 놓았습니다}) \]

- ‘개’ is also ambiguous between ECS and EAU. When it precedes an auxiliary verb, it is EAU; it is ECS in all other environments.

\[ 먹 개 둡였습니다: \text{먹/VV+개/EAU 둡/VX+없/EPF+먹/EFN} \]
\[ 먹 개 둡았습니다: \text{먹/VV+개/ECS 둡/VV+없/EPF+먹/EFN} \]
\[ 느릿 개 달렸습니다: \text{느릿/VJ+개/ECS 달/VV+없/EPF+먹/EFN} \]
4.13 ENM or EAU

- ‘기’ is not always tagged as a nominal ending (ENM). In sentences that have a verb or an adjective suffixed with ‘기는’, ‘기만’, or ‘기로’ followed by an auxiliary verb, ‘기는’, ‘기만’, or ‘기로’ is tagged as ECS. Do not give ENM tag to ‘기’ and one of the postposition tags to ‘는, 만, 로’ in such sentences.

손 먹은 밥을 먹/VV+기는/ECS 후/VX+었/EPF+ démarch/EFN.
날씨가 열/VJ+기만/ECS 후/VX+었/EFN.
우리는 연규소의 시설을 사용/NNC+하/XSV+기로/ECS 후/VX+ démarch/EAU 었다.

4.14 EFN or NNX

- Be careful to distinguish between the two different usages of ‘ㄴ 지’. When it is used to end an interrogative clause, it is a final ending (EFN):

철수가 누구 인지는 모르겠 말:  누구/NPN+의/CO+ㄴ 지/EFN+은/PAU

- However, in the following example, ‘ㄴ 지’ is roughly translatable as ‘since the time when ...’. In this case, it in fact consists of two parts, ‘지/NNX’ as a dependent noun, and ‘ㄴ/EAN’ which is the verbal ending part of the relative clause modifying it.

학교에 간 지가 오래 됐 말:  간/VV+ㄴ/EAN 지/NNX+같/PCA

5 Confusing Examples

- ‘그럼’
  If it means ‘yes’, it is tagged as IJ. If it means ‘if so’, it is tagged as ADC.

- ‘때로’
  If ‘때로’ immediately follows a noun without a space, it is a postposition marker and therefore is tagged as PAD.

명령 때로: 명령/NNC+때로/PAD

If ‘때로’ is modified by a relative clause, it is a dependent noun (NNX).

말한 때로: 말한 때로/NNX

- ‘(이)หลายคน’
  If ‘(이)หลายคน’ is attached to a noun which is a complement of a verb, it should be tagged as PAD as a whole.
In some cases, ‘(어) 맘고’ can be used to form an echo question, in which a noun that has been mentioned by the speaker is questioned back by the hearer. In this case, noun+‘(어) 맘고’ is not a complete sentence but rather a sentence fragment; therefore there is no need to treat ‘(어) 맘고’ as a sentence final ending (EFN) and is tagged as PAD.

A: 그는 부산에서 삐니까.
B: 부산/NPR+여 맘고/PAD ?/SFN

• ‘말라’
In most cases, ‘말’ in ‘말라’ is an auxiliary verb and is tagged as VX. But it can also be used as a main verb, in which case it is tagged as VV.

가/VV+직/EAU 맘/VX+ 맘/EFN
너 말/VV+고/ECS 네가 맘Collision

• ‘모르다’
In general, ‘모르다’ is tagged as VV. But when it follows a predicate and means ‘might’, it is tagged as VX:

손이가 노래/NNC+악/XXSF+ㄹ 진도/ECS 모르/VX+ㄹ 맘/EFN.

• ‘밖에’
If ‘밖에’ means ‘only’, it is tagged as PAU as a whole. If ‘밖에’ means ‘outside’ , it is tagged as NNC. If it follows a demonstrative and means ‘besides’, it should be tagged as NNX.

너밖에 없어: 너/NPR+밖에/PAU 없어
장문 밖에: 장문/NNC 밖/NNC+예/PAK
그 밖에 무엇을 알까요: 그/DAN 밖/NNX+예/PAD

• ‘별’
‘별’ can be tagged as XSF or DAN.

별/NPX+ 별/XXSF
별 이상한 소리: 별/DAN 이상한 소리

• ‘본’
‘이본, 그본, 저본’ are tagged as NPN. But ‘본’ can be a separate word, and in this case, it should be tagged as NNX.
이 명사에 붙는 '본'은 'original'의 의미를 가질 때 사용됩니다. 

본/XP+책/N
본 밖으로: 본/XP+책/N

이 명사에 붙는 '상'은 'appearance'의 의미를 가릴 때 사용합니다. 

상/NC
상 밖으로: 상/NC+에/P

이 명사에 붙는 '소'는 'place'의 의미를 가릴 때 사용됩니다. 

소/N
소 밖으로: 소/XP+에/N

이 명사에 붙는 '수'는 '수'로 본은 preceding noun의 prefix입니다. 

수/XP+표/N
수 밖으로: 수/DAN 표/N

이 명사에 붙는 '생'은 'calculation'의 의미를 가질 때 사용됩니다. 

생/NC
생 밖으로: 생/XP+표/N

32
- '시간'
If it is used as a unit meaning 'hour', it is tagged as NNX.

한 시간/NNX, 두 시간/NNX

Otherwise, it is tagged as NNC.

시간/NNC+의 없탁.
공부하는 시간/NNC

- '오'

'오' is tagged as PAU if it attaches to a noun, but it is tagged as EFN if it attaches to a verb::

'나는 오' : 나/NPN+는/PAU+오/PAU
'있어 오' : 있/EFF+있는/EFN

- '이'
The copula '이' does not surface when the noun it is attached to ends in a vowel. There are a couple of kinds of contexts where special care should be taken to make sure it is recovered:

학교 마다 : 학교/NNC+의/CO+있다/EFN

In many contexts where 'noun+있다' occurs, '이' must be identified between the noun and '있다'. Adnominal ending '있다' presents one such case:

완달락는 물고기 : 완달/NPR+있는/EAN 물고기
철수가 완달락는 사실 : 철수가 완달/NPR+있는/EAN 사실

Contrast the above example with '있다' with seemingly similar examples with '있다' below. Copula '이' is identified/recovered only when '있다' attaches to a predicative noun in a clause, and not when it attaches to an argument noun. In '완달락고 부분탁', '있다' is tagged as PAD, enabling us to treat '완달락고' as the internal argument of the verb '부탁'. On the other hand, in '나는 철수가 완달락고 망는 맛', '철수가 완달락고 맛' forms a clause and therefore a predicate head is needed; copula '이' is supplied for this reason.
Moreover, ‘탁고’ is tagged as ECS in the context where copula ‘의’ is recovered, while it is tagged as PAD if it is on an argument noun. Note that ‘의’ in ‘의탁고’ on an argument noun is not a copula, but is merely an epenthetic vowel. And so, ‘의탁고’ in this case is tagged as PAD.

In most cases, ‘의－’ is a combination of a copula and a sentence final ending marker. But when it attaches to nouns in a list context, it should be tagged as PAD as a whole.

When ‘지’ occurs after an adjective, it signifies change of state. In this case it is tagged as VX.

In contrast, when ‘지’ occurs on a verb, it functions as a passivizer. It does not get a separate tag in this case, and ‘verb+지’ as a whole are tagged as VV.

Although it is considered ungrammatical according to some prescriptive Korean grammars, ‘지’ may occur with verbs that have already been passivized. In such cases, ‘지’ is treated as an auxiliary verb and tagged as VX, not VV.

‘폭’ ‘등폭, 서폭, 남폭, 동폭, ..., ’ are tagged as NNC as a whole. But ‘폭’ can be a separate word and it is tagged as NNX in this case.
학교/NNC 종/NNX
신생/NNC 종/NNX

- 측
측’ is a suffix. But it can also be a separate word and it is tagged as NNX in this case.
주석/NNC 측/XSF
우연/NNP 측/NNX의/PCA 의견/NNC

But ‘측, 우측’ should be tagged as NNC as a whole.

- 타
타’ is a prefix. But in some very restricted cases when it occurs as a separate word, it is tagged as NNC.

- 하다
If ‘하다’ occurs after a verb or an adjective and means ‘must’ or ‘try to, be about to’, it is tagged as VX.

- 하다’ can occur after a verb or an adjective that has endings such as ‘기도’, ‘기만’, ‘기는’, or ‘기토’. In such cases, ‘하다’ should also be tagged as VX. Note that the tag for the endings on the preceding verbs is ECS.

On the other hand, if ‘하다’ is used as a causative verb, it is a main verb and therefore is tagged as VV.

- 하다
If ‘하다’ is used as a light verb, it is tagged as VV.
If ‘하다’ is used as a light adjective, it is tagged as VJ.

손어가 흔들/NNC+은/PAU 하/VJ+тся/EFN

If ‘하다’ is used as a pro-verb meaning ‘say’, it is tagged as VV.

손어는 철수가 박보/NNC+이/CO+학교/ECS 하/VV+었/EPF+께서/EFN

• ‘단’

Be careful to distinguish between ‘단/NNU’ and ‘단/DAN’. The former means ‘one’, whereas the latter means ‘about’.

단/NNU 사람만 갔다.
단/DAN 2000 미터 정도 됐다.

• ‘모’

‘모’ in ‘상리모’ should not be separated out, and the entire lexical item should be tagged as NNC. This is because ‘상리모’ is a name of a type of an armed vehicle. But if ‘모’ is used as a unit, it should be tagged as NNX.

모/NNU 모/NNX