



University of Pennsylvania
ScholarlyCommons

IRCS Technical Reports Series

Institute for Research in Cognitive Science

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

University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-01-09.

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

University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-01-09.

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	Adverbial 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, ‘**절** **데**’ 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

Category	Tag Description	Tag Label	Examples
noun	proper noun	NPR	한국 (Korea), 클린톤 (Clinton)
	common noun	NNC	학교 (school), 컴퓨터 (computer)
	dependent noun	NNX	것 (thing), 등 (etc), 년 (year), 달러 (dollar), 적 (situation)
	personal pronoun, demonstrative pro- noun	NPN	그 (he), 이것 (this), 무엇 (what)
	ordinal, cardinal, nu- meral	NNU	하나 (one), 첫째 (first), 1, 세 (three)
	words written in for- eign characters	NFW	Clinton, computer
predicate	verb	VV	가 (go), 먹 (eat)
	adjective	VJ	예쁘 (pretty), 다르 (different)
	auxiliary predicate	VX	있 (present progressive), 야 (must)
adverb	constituent adverb, clausal adverb	ADV	매우 (very), 조용히 (quietly), 제발 (please), 만일 (if)
	conjunctive adverb	ADC	그리고 (and), 그러나 (but, however), 및 (and), 혹은 (or)
adnominal	configurative, demon- strative	DAN	새 (new), 헌 (old), 그 (that)
interjection	exclamation	IJ	아 (ah)
list	list marker	LST	a, (b), 1, 2.3.1, 가, 나.

2.2 Function Tags

Category	Tag Description	Tag Label	Examples
postposition	case	PCA	가/이 (nominative), 을/를 (accusative), 의 (possessive), 야 (vocative)
	adverbial	PAD	에서 (from), 로 (to)
	conjunctive	PCJ	와/과, 하고 (and)
	auxiliary	PAU	만 (only), 도 (also), 는 (topic), 마저 (even)
copula		CO	이 (be)
ending	final	EFN	는 다/ㄴ 다 (declarative), 니, 는 가, 는 지 (interrogative), 어 락/랴 (imperative), 자 (propositive), 구 나 (exclamatory)
	coordinate, subordinate, adverbial, complementizer	ECS	고 (and), 므로 (because), 게 (attaches to adjectives to derive adverbs), 다고 (that), 략고 (that)
	auxiliary	EAU	아, 게, 지, 고 (on verbs or adjectives that immediately precede auxiliary predicates)
	adnominal	EAN	는/ㄴ (on main verbs or adjectives in relative clauses or complement clauses of a complex NP)
	nominal	ENM	기, 음 (on nominalized verb)
	pre-final ending (tense, honorific)	EPF	었 (past), 시 (honorific), 겠 (future)
affix	suffix	XSF	님, 들, 적
	prefix	XPF	제, 각, 매
	verbalization suffix	XSV	아, 되, 시 키
	adjectivization suffix	XSJ	스럽, 답, 아

2.3 Symbols

Category	Tag Description	Tag Label	Examples
comma		SCM	,
termination	sentence ending markers	SFN	. ? !
left quotation mark		SLQ	‘ (“ {
right quotation mark		SRQ	’) ” }
symbols	others	SSY	... ; : -

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)

- * 한국들
Korea-Plural
- * 홍길동들
Hongkiltong-Plural

(2) Common nouns (NNC), personal pronouns (NPN)

- 건물들
building-Plural
- 그들
he-Plural

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

클린톤/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, 자전거, 오리, 사람, 동물, 장관, 할아버지, 언어학, 나무, 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).

수영/NNC+ㅅ/CSV+다
꼭딱/NNC+ㅅ/CSV+다
노태/NNC+ㅅ/CSV+다

여성/NNC+ㅅ/CSJ+다
여성/NNC+ㅅ/CSJ+다
필요/NNC+ㅅ/CSJ+다

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

에르-106/NNC
컴퓨터/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).

에르-106/NNC
승티-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

- ‘하루, 이틀, 사흘, ...’ are tagged as NNC.

- ‘이후, 그후, 이전, 그전’ and ‘이때, 그때, 저때, 이번, 저번’ are each treated as one word and are tagged as NNC.

In the corpus, there may be a space in between ‘이, 그, 저’ and ‘때, 번’. In this case, ‘이, 그, 저’ should be tagged as DAN and ‘때’ should be tagged as NNC and ‘번’ should be tagged as NNX.

이 때: 이/DAN 때/NNC
 저 번: 저/DAN 번/NNX

- ‘그날, 이날, 오늘날, 전날, 날’ are tagged as NNC as a whole.
- ‘때’ is tagged as NNC, not NNX, because it can be used as an independent noun.

때가 때이니 만큼...

- ‘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, 전체’ are tagged as NNC; they are not dependent nouns.

사이/NNC+가/PCA 벌어졌다
 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 비행기’ 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.

- ‘산하’ is tagged as NNC.

109/NNU 연태/NNC 산하/NNC

- ‘기탁’ is tagged as NNC.

기탁/NNC 사망/NNC

- ‘편’ (different team), ‘면’ (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:

무렵, 쯤, 적, 곳, 줄, 뿐, 등, 수, 한, 지, 때문, 대로

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

미터, 달락, 개, 번, 월, 일, 년, 달, 주, 주일, 권, 명, 마리, 시간, 킬로미터, 센티, 가지, 군데

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

년/NNC+간/XSF
월/NNC 2 회
일/NNC 3 회

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

달락/NNC 가치/NNC
미터/NNC 제도/NNC

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

제2 차: 제/XPF+2/NNU 차/NNX
제2 회: 제/XPF+2/NNU 회/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
평화/NNC 시/NNX

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

- ‘것, 뿐’ in ‘것이 다’ and ‘뿐이 다’ are NNX.

- When the dependent noun ‘것, 터’ takes other bound morphemes, phonological contraction tends to take place. In this case, ‘것, 터’ and the morphemes should be separated out and tagged accordingly:

건 가: 것 /NNX+ 이 /CO+ㄴ 가 /EFN
 계: 것 /NNX+ 이 /PCA
 있을 텐 데: 터 /NNX+ 이 /CO+ㄴ 데 /EFN

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

한 번 /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 주 일 /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.

말씀 드린 박 /NNX+와 /PAD 같이 /ADV

3.4 Pronoun (NPN)

- Pronouns include personal pronouns, interrogative-indefinites (so-called ‘*wh*-words’) and demonstrative pronouns.
- ‘이것, 그것, 저것, 여기, 저기, 이곳, 그곳, 저곳’ are each treated as one word and tagged as pronouns (NPN).
- Some interrogative-indefinites are tagged as pronouns (NPN):

누구, 무엇, 언제, 어디, 얼마, 아무

- ‘누구나, 무엇이냐, 아무나’ are tagged as NPN as a whole.

- ‘왜, 얼마나, 얼마든지, 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, 귀학’ are polite second person pronoun forms; they are tagged as NPN.
- ‘자기’ is NPN.
- When ‘어딘가/어디인가, 누군가/누구인가’ correspond to ‘somewhere, someone’ in meaning, they are recognized as single words and should be tagged as NPN as a whole:

어딘가로 갔다: 어딘가/NPN+로/PAD
 누구인가를 만났다: 누구인가/NPN+를/PCA

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

자네 집이 어디가?: 어디/NPN+이/CO+ㄴ가/EFN
 저 사람이 누구인가 모르겠다: 누구/NPN+이/CO+ㄴ가/EFN

3.5 Number (NNU)

- Number words (NNU) include cardinals (‘하나, 둘, 한, 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: ‘하나’ and ‘한’ both receive the NNU tag since they are numeric in terms of meaning, although syntactically they are quite different in that ‘하나’ functions as a noun and ‘한’ as a prenominal modifier.
- ‘두어, 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 ‘며칠’ 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 ‘클린톤’ is tagged as NPR.
- In cases like ‘R(에르)-116’ and ‘HAK(하크)-24,’ erase the material in parenthesis and tag them as NFW.

R(에르)-116: R-116/NFW

Hak(하크)-24: H-24/NFW

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

mm/NFW

km/NFW

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/니’, adjectives cannot:

- (3) a. 철수가 뛰다.
Chelswu-Nom run-Pres-Decl
‘Chelswu is running.’
- b. *꽃이 예쁘다.
flower-Nom pretty-Pres-Decl
‘The flower is pretty.’

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

- (4) a. 철수가 뛰고 있다.
Chelswu-Nom run be-Decl
‘Chelswu is running.’
- b. *꽃이 예쁘고 있다.
flower-Nom pretty be-Decl
‘The flower is being pretty.’

- ‘있다, 없다, 계시다, 아니다’ are tagged as adjectives (VJ).
- If ‘있다, 계시다’ are used as auxiliary predicates, then they should be tagged as VX.

먹/VV+고/EAU 있/VX+다/EFN
 잡수시/VV+고/EAU 계시/VX+다/EFN

3.9 Auxiliary Predicate (VX)

- Auxiliary predicates include auxiliary verbs such as ‘놓다, 되다’ as well as auxiliary adjectives such as ‘있다, 싶다’. They cannot stand alone without a preceding verb or an adjective. Some examples of auxiliary predicates include:

놓다, 있다, 되다, 나다, 오다, 버리다

바꾸/VV+어/EAU 놓/VX+지요/EFN
 쉬/VV+고/EAU 나/VX+서/PCS
 먹/VV+게/EAU 되/VX+었/EPF+다/EFN
 밝/VJ+아/EAU 오/VX+았/EPF+다/EFN
 깨/VV+어/EAU 버리/VX+다/EFN

싶다, 보다, 아니하다

노태/NNC+하/XSV+고/EAU 싶/VX+다/EFN
 노태/NNC+하/XSV+는가/EFN 싶/VX+다/EFN
 찢/VV+어/EAU 보/VX+았/EPF+다/EFN
 찢/VV+는가/EFN 보/VX+다/EFN
 나쁘/VJ+지/EAU 앎/VX+다/EFN
 먹/VV+지/EAU 앎/VX+았/EPF+다/EFN

만하다, 듯하다, 법하다, 적하다, 뻔하다

볼 만하다: 보/VV+르/EAN 만하/VX+다/EFN
 우는 듯하다: 울/VV+는/EAN 듯하/VX+다/EFN
 착할 법하다: 착하/VJ+르/EAN 법하/VX+다/EFN
 아픈 적하다: 아프/VJ+ㄴ/EAN 적하/VX+다/EFN
 다칠 뻔했다: 다치/VV+르/EAN 뻔하/VX+였/EPF+다/EFN

Sometimes ‘만, 듯, 법, 적, 뻔’ accompany a postposition and are separated from the verbal stem ‘하’; in these cases they must be tagged as NNX:

볼 만도 하다: 보/VV+르/EAN 만/NNX+도/PAU 하/VV+다/EFN

을 듯도 아다: 우 /VV+ㄹ /EAN 듯 /NNX+도 /PAU 학 /VV+ 다 /EFN
 작할 법은 아다: 작학 /VJ+ㄹ /EAN 법 /NNX+은 /PAU 학 /VV+ 다 /EFN
 아픈 척을 아다: 아프 /VJ+ㄴ /EAN 척 /NNX+을 /PCA 학 /VV+ 다 /EFN
 다칠 뻔도 아다: 다치 /VV+ㄹ /EAN 뻔 /NNX+도 /PAU 학 /VV+ 다 /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.

먹어 보았다: 먹 /VV+어 /EAU 보 /VX+았 /EPF+ 다 /EFN
 텔레비전을 보았다: 텔레비 /NNC+를 /PCA 보 /VV+ 았 /EPF+ 다 /EFN

빵을 버렸다: 빵을 버리 /VV+었 /EPF+ 다 /EFN
 빵을 먹어 버렸다: 빵을 먹 /VV+어 /EAU 버리 /VX+었 /EPF+ 다 /EFN

3.10 Verbal Adverb, Clausal Adverb (ADV)

- Adverbs modify either the entire clause (clausal adverbs) or the verb phrase (phrasal adverbs). Clausal adverbs include ‘아마, 제발, 혹시, 과연, 만일, 만약, 비록, 부디, 설마, 다행히, 다만, 또한, 또’.

Phrasal adverbs describe manner, time, place or degree. They include ‘빨리, 천천히, 바로, 자주, 벌써, 아까, 즉시, 매우’.

Onomatopoeic manner adverbs or phenomimes include ‘소곤소곤, 딸각딸각, 딸랑딸랑, 아장아장, 반짝반짝, 싱글빙글’.

Some other examples include:

오래, 모두, 보통, 통상, 직접, 즉시, 얼마든지, 얼마나, 얼마만큼, 어떻게, 이렇게, 그렇게, 저렇게, 역시, 거꾸로, 항상, 전혀, 이따금, 이따금씩, 가끔, 가끔씩, 대부분, 대략, 다, 각각

- ‘그리, 그리로, 저리, 저리로’ are tagged as ADV.
- Adverbs derived from an adjective with suffix ‘이’ are tagged as ADV as a whole, without breaking them into the adjective and the suffix.

없이 /ADV
 같이 /ADV

- ‘기껏해야’ is a single word and is tagged as ADV.

- ‘또, 또한’ are tagged as ADV, not ADC. But ‘또는’ 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 또 시험에 떨어졌다.
 철수 또한 가야 한다.
 점심 또는 저녁

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

안 먹었다: 안/ADV 먹/VV+었/EPF+다/EFN
 못 먹었다: 못/ADV 먹/VV+었/EPF+다/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, 그러나, 그러니까, 그러면, 그러므로, 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 **D**eterminer-**A**d**N**ominal. Examples are:

이, 그, 저, 오, 새, 헌, 여러, 약, 한, 무슨, 어떤, 이런, 저런, 그런, 어느, 모든, 아무, 각, 전, 제, 주, 지난, 별, 다른, 바로

Note that some of these examples have ambiguous usages. For example, **한** 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’.

전 군 : 전/DAN 군/NNC
 전 연 대 : 전/DAN 연 대/NNC
 주 전 투 전 차 : 주/DAN 전 투/NNC 전 차/NNC

3.13 Interjection (IJ)

- ‘예’ should be tagged as IJ. But ‘아니오 , 아닙니다’ are viewed as inflected forms of the adjective ‘아니’ and therefore should be given VJ tag plus appropriate tags for the inflections.

아 니/VJ+ 오 /EFN
 아 니/VJ+ 떠 니 다/EFN

- ‘글썄,’ and ‘글썄요’ 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
 가./LST
 나./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 (‘가/이’), accusative marker (‘을/를’), genitive marker (‘의’), and vocative marker (‘야’).

- The marker ‘가/이’ 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).

내가 팔이 마비됐다:

나/NPN+가/PCA 팔/NNC+이/PCA 마비/NNC+되/XSV+있/EPF+다

3.16 Adverbial Postposition (PAD)

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

에게, 으로, 까지, 부터, 한테, 로부터, 예다가, 예서부터

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

- Comparative markers such as ‘처럼, 보다’ are also tagged as PAD.
- ‘으로’ attached to ‘Noun-적’ should be tagged as PAD.

단 계/NNC+적/XSF+으로/PAD

- ‘라고’ attached to a noun should be tagged as PAD.

완다/NPR+라고/PAD 부른다.

However, the same ‘라고’ 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 완다가 물고기라고 믿는다.

완다/NPR+가/PCA 물고기/NNC+이/CO+라고/ECS

But seemingly similar ‘라는’ 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 ‘라는’. Refer to the entries for ‘라고’ and ‘이’ in §5 for more information.

완다라는 물고기

완다/NPR+이/CO+라는/EAN 물고기/NNC

- ‘한테’ and ‘한테서’ 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’.

수리병한테: 수리병/NNC+한테/PAD

수리병한테서: 수리병/NNC+한테서/PAD

3.17 Conjunctive Postposition (PCJ)

- Conjunctive postpositions correspond to English ‘and’, or ‘or’. Some examples include:

하고, 나/이나, 와/과, 락든 지/이 락든 지

- 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, 밖 에, 마 따, 조 차, 까 지, 부 터’.

Note that some of these examples (e.g., 까 지, 부 터) are ambiguous between auxiliary postposition usage and adverbial postposition usage. See §4.6 for ways to distinguish between the two usages.

- ‘(이) 란’ which is short for ‘라고 하는 것은’ is tagged as PAU.

간 단한 수 리/NNC+ 란/PAU 어 떤 종 류의 수 리를 말 하는 가?
근 의 반/NNC+ 이 란/PAU 무 엇 인 가?

- The postposition ‘나, 이 나’ should be tagged as PAU:

몇 도 나: 몇/NNU 도/NNX+ 나/PAU
세 번 이 나: 세/NNU 번/NNX+ 이 나/PAU

Note that ‘나, 이 나’ can also be used as a conjunctive postposition. In this case, they should be tagged PCJ. See §4.8 for how to distinguish PCJ and PAU usages for ‘나, 이 나’.

- Complex postpositions that have undergone phonological contraction should be separated out into individual parts and tagged accordingly:

대 대 언: 대 대/NNC+ 예/PAD+ 는/PAU
동 무 예 겠: 동 무/NNC+ 예 게/PAD+ 는/PAU
겨 울 철 언: 겨 울 철/NNC+ 예/+PAD+ 는/PAU

3.19 Copula (CO)

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

박보이 닻: 박보/NNC+이/CO+ 닻/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 ‘이’ 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 봅 닻
-는 가/EFN 싶 닻

On verbs in complement clauses:

-는 가/EFN+예/PAD 달렸 닻
-는 지/EFN+를/PCA 모르 겠 닻
-아 지 않/VX+았/EPF+나/EFN 생각한 닻

- ‘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, 냐, 닻’ in terms of its syntactic distribution; therefore it is tagged as EFN, not ECS.

철수는 가는 지?
철수는 가는 지요?

가는 가 말아라:	가/VV+는 가/EFN 말아라
가는 지 말아라:	가/VV+는 지/EFN 말아라
가는 가 싶 닻:	가/VV+는 가/EFN 싶 닻
가는 지 싶 닻:	가/VV+는 지/EFN 싶 닻
가는 가를 말아라:	가/VV+는 가/EFN+를/PCA 말아라
가는 지를 말아라:	가/VV+는 지/EFN+를/PCA 말아라
가는 냐가 문제이다:	가/VV+는 냐/EFN+가/PCA 문제이다
가는 지가 문제이다:	가/VV+는 지/EFN+가/PCA 문제이다

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 ‘말아다, 생각아다, 믿다, 요구아다’ are tagged as ECS. Examples include ‘다고’ and ‘라고’. See §5 for further information on the item ‘라고’.

그 곳에 전파기가 있다고 생각한다: 있/VJ+다고/ECS
 편지를 쓴다고 말했다: 쓰/VV+니다고/ECS

- Conditional markers ‘다면’ and ‘라면’ are tagged as ECS:

있다면: 있/VJ+다면/ECS
 가라면: 가/VV+라면/ECS

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

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

- Ending ‘는 데’ must be tagged as ECS.

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

철수는 있/VJ+는 데/ECS 영희는 없다. (meaning ‘있지만’)
 꽃이 있/VJ+는/EAN 데/NNX 나비가 있다. (meaning ‘있는 곳에’)

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

설치/NNC+아/XSV+도록/ECS 뒤/VX+어/EAU 있/VX+다/EFN

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

일찍 끝내도록 노력해라.

Endings ‘어야’, and ‘어도’ on verbs or adjectives that precede modal auxiliary predicates are tagged as ECS.

설치하여야 한다: 설치/NNC+하/XSV+어야/ECS 하/VX+ㄴ /EPF+다/EFN

설치하여도 된다: 설치/NNC+하/XSV+어도/ECS 하/VX+ㄴ /EPF+다/EFN

‘어야’, and ‘어도’ are ECS since they can occur on main verbs in subordinate clauses.

에어콘을 설치하여야 방이 시원하다.

에어콘을 설치하여도 방이 덥다.

3.22 Auxiliary Ending (EAU)

- Auxiliary endings occur on verbs or adjectives that are followed by auxiliary predicates. Auxiliary endings include ‘아,’ ‘게,’ ‘지,’ ‘고,’ their morpho-/phonological variations and nothing else:

먹어 보았다: 먹/VV+어/EAU 보/VX+았/EPF+다/EFN

먹게 되었다: 먹/VV+게/EAU 되/VX+었/EPF+다/EFN

먹지 않았다: 먹/VV+지/EAU 았/VX+았/EPF+다/EFN

먹고 있다: 먹/VV+고/EAU 있/VX+다/EFN

고장나 버렸다: 고장나/VV+아/EAU 버티/VX+었/EPF+다/EFN

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

쓸 만하다: 쓰/VV+ㄴ/EAN 만하/VX+다/EFN

통할 만한: 통하/VV+ㄴ/EAN 만하/VX+ㄴ/EAN

갈 법하다: 가/VV+ㄴ/EAN 법하/VX+다/EFN

Also, some auxiliary verbs take a verbal complement with an ECS ending. Examples are:

먹/VV+어야/ECS 하/VX+ㄴ 다/EFN.

가/VV+어도/ECS 되/VX+ㄴ 다/EFN.

3.23 Adnominal Ending (EAN)

- Adnominal endings ‘는, ㄴ, 을, ㄴ, ㄴ’ occur on predicates in a relative clause or complement clause of a complex NP.
- In cases where ‘ㄴ 다는’ is a short for ‘ㄴ 다고 하는’, ‘ㄴ 다는’ should simply be tagged as EAN.

과열된다는 보고서: 과열/NNC+되/XSV+ㄴ 다는/EAN 보고서/NNC

- ‘라는’ as in ‘학생이라는’ is tagged as EAN. Note the presence of the copula ‘이’ between the noun and ‘라는’. 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.

철수가 학생이라는 사실: 철수가 학생/NNC+이/CO+라는/EAN 사실/NNC

완다라는 물고기: 완다/NPR+이/CO+라는/EAN 물고기/NNC

3.24 Nominal Ending (ENM)

- Nominal endings ‘기, 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:

파괴함으로써: 파괴/NNC+하/XSV+□/ENM+으로써/PAD

3.25 Pre-Final Ending (EPF)

- Pre-Final endings include tense and honorific markers, such as ‘었/았, 겠, 시/으시’.
- These morphemes can co-occur, as in ‘셨, 었었, 었겠, 션겠, 시곶’. They are not further analyzed, and receive a EPF tag as a whole:

가셨곶다: 가/VV+셨곶/EPF+다/EFN
 먹었었다: 먹/VV+었었/EPF+다/EFN

3.26 Suffix (XSF)

- Suffixes mostly attach to nouns. Some examples include:

님, 네, 네들, 네들끼리, 쯤, 깨, 씨, 기리, 식, 째, 여, 용, 간, 당, 들간, 들끼리, 들쯤, 상, 짜리, 식, 확, 급, 실

선생님: 선생/NNC+님/XSF
 도의상: 도의/NNC+상/XSF
 이론상: 이론/NNC+상/XSF
 무전기용: 무전기/NNC+용/XSF
 하나씩: 하나/NNU+씩/XSF
 자기네: 자기/NPN+네/XSF
 10미터짜리: 10/NNU 미터/NNX+짜리/XSF
 암호확: 암호/NNC+확/XSF
 현대식: 현대/NNC+식/XSF
 지휘관실: 지휘관/NNC+실/XSF
 시간당: 시간/NNC+당/XSF

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

아이/NNC+들/XSF
 빨리/ADV+들/XSF 먹어라

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.

단 계/NNC+적/XSF

사 회/NNC+적/XSF

But note that ‘비교적’ 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.

비 교 적/ADV

- ‘구, 시, 군, 면, 읍, 리, 동’ occur with city, town, place names. These should be tagged as XSF.
- ‘성, 장, 제, 품, 망, 소, 표, 호, 부, 병, 권’ and ‘대’ are not recognized as suffixes. Nouns occurring with these morphemes should simply be tagged as NNC as a whole.

특 무 장/NNC

구 분 대/NNC

가 능 성/NNC

소 련 제/NNC

부 분 품/NNC

통 신 망/NNC

감 시 소/NNC

지 휘 소/NNC

승 리-58 호/NNC

암 오 표/NNC

포 병 부/NNC

경 보 병/NNC

지 휘 권/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’.

삼 선/NNC, 오 선/NNC 철 조 망

방 사 기 조/NNC

대 기 조/NNC

3.27 Verbalization Suffix (XSV)

- Verbalization suffixes attach to nouns to turn them into a verb. They include ‘**학**, **되**, **시키**, **발**.’
- Verbs that are derived by a verbalization suffix should be separated into the noun stem and the suffix part and tagged accordingly.

수영/NNC+학/XSV+다/EFN
폭박/NNC+되/XSV+다/EFN
노태/NNC+시키/XSV+다/EFN

- But the following items are viewed as frozen forms, and therefore ‘**학**’ should not be separated out:

말하다: 말학/VV+다/EFN
가하다: 가학/VV+다/EFN
못하다: 못학/VX+다/EFN
만하다: 만학/VX+다/EFN
듯하다: 듯학/VX+다/EFN

- ‘**나**’ 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 ‘**나**’ should be tagged as a verb.

고장나다: 고장나/VV+다/EFN
바람나다: 바람나/VV+다/EFN
불나다: 불나/VV+다/EFN

고장 나다: 고장/NNC 나/VV+다/EFN
바람 나다: 바람/NNC 나/VV+다/EFN
불 나다: 불/NNC 나/VV+다

- ‘**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 ‘**말씀드리다**, **인사드리다**’ are recognized as a word as a whole.

말씀드리다: 말씀드리/VV+다/EFN
말씀 드리다: 말씀/NNC 드리/VV+다/EFN
선물드리다: 선물/NNC 드리/VV+다/EFN

- ‘**발**’, 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.

보충발다: 보충/NNC 발/VV+다/EFN
보충발다: 보충/NNC+발/XSV+다/EFN

3.28 Adjectivization Suffix (XSJ)

- Adjectivization suffixes attach to nouns to turn them into adjectives. They include ‘스럽, 답, 학.’
- Adjectives that are derived by an adjectivization suffix should be separated into the stem and the suffix, and tagged accordingly.

여성/NNC+스럽/XSJ+타/EFN
여성/NNC+답/XSJ+타/EFN
투명/NNC+학/XSJ+타/EFN

Note that ‘학’ 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:

제, 각, 반, 비

제23: 제/XPF+23/NNU
각급: 각/XPF+급/NNC
반공산주의: 반/XPF+공산주의/NNC
비정규전: 비/XPF+정규전/NNC

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.

매시간: 매/XPF+시간/NNC
매학기: 매/XPF+학기/NNC

But ‘매일’ is tagged as NNC as a whole.

- ‘총’ is not tagged as XPF:

총검열/NNC

- In general, when ‘대, 소’ precede a noun, they are tagged as XPF. But some frozen words such as given below are tagged as NNC as a whole.

대규모 /NNC
소규모 /NNC
대학교 /NNC
중학교 /NNC
소책자 /NNC
대책자 /NNC

- In general, ‘재’ (again) is tagged as XPF. But some frozen words that are prefixed with ‘재’ are tagged as NNC as a whole.

재전개: 재/XPF+전개/NNC
재결합: 재결합/NNC
재교육: 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 ‘피’ is not tagged as XPF because it is part of the domain-specific word ‘피지원’.

피지원 /NNC
반탱크 /NNC
반항공 /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 (오제엠-3):
OZM-3/NFW -LRB-/SLQ 오제엠-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, 실제, 만약, 만일, 절대, 아까, 매일, 즉시, 도합, 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 좋았다.
절대/ADV 안된다.
만약/ADV 간단
아까/ADV 갔다.
그동안/ADV 고맙았다.

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

진짜/NNC+가/PCA 아니야.
정말/NNC+이/CO+약/EFN
만약/NNC+에/PAD
만일/NNC+의/PCA 경우/NNC
절대/NNC 온도/NNC
아까/NNC+의/PCA 약속
매일/NNC+의/PCA 일과
그 즉시/NNC+에/PAD+는/PAU 제 정신이 아니었다
그동안/NNC+의/PCA 실적/NNC

Also note that ‘정말로, 진짜로, 실제로, 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.

아이들은 모두/ADV 모여라.
모두/NPN+가/PCA 왔다.

4.3 NPN or DAN

- Some words such as ‘이, 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.

이/NPN+는/PAU 무엇인가
이/DAN 사람

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.’

똑같은 식/NNX+으로/PAD

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

재택/NNC+식/XSF
한국/NNC+식/XSF

But note that in ‘변조식’, ‘식’ does not get a separate tag and the entire lexical item is tagged as NNC, as ‘변조식’ is recognized as a domain-specific lexical item.

변조식/NNC

If ‘식’ means ‘equation’, it should be tagged as NNC.

- ‘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.

여자/NNC+들/XSF
아이/NNC+들/XSF
빨리/ADV+들/XSF 먹어라

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

소, 개, 닭들/NNX

- 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.

감시소와 지휘소/NNC+간/XSF+예/PAD
이러나 저러나 간/NNX+예/PAD
십여년 간/NNX+예/PAD

4.5 VJ or DAN

- Note that ‘다른’ 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 ‘다르다’ and therefore should be tagged as 다르/VJ+ㄴ/EAN.

다른/DAN 사람들
나와는 참 다르/VJ+ㄴ/EAN 사람들

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.

너부터 먹어라. (PAU)
너까지 나한테 이럴 수가 있니? (PAU)
나는 기차역까지 걸어 갔다. (PAD)
나는 학교에서부터 기차역까지 걸어 갔다. (PAD)

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 ‘같다, 다르다’, it is an adverbial postposition, and is tagged as PAD.

(7) Conjunctive postposition (PCJ)

- a. 너와 나는 다르다.
‘You and I are different.’
- b. 너하고 나는 같다.
‘You and I are same.’

(8) Adverbial postposition (PAD)

- a. 나는 너와 다르다.
‘I am different from you.’
- b. 나는 너하고 같다.
‘I am same as you.’

4.8 PCJ or PAU

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

사과나 오렌지를 먹었다: 사과/NNC+나/PCJ 오렌지/NNC+를/PCA
세 번이나 읽었다: 세/NNU 번/NNX+이 나/PAU

4.9 XSF or NNC

- If ‘과’ or ‘급’ 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.

영문과: 영문/NNC+과/XSF
좋은 과: 좋은 과/NNC
각과: 각/XPF+과/NNC

중대급: 중대/NNC+급/XSF
급이 다르다: 급/NNC+이/PCA 다르다
각급: 각/XPF+급/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

- ‘바로’
‘바로’ 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.

바로/ADV 학교로 갔다.
바로/DAN 저/DAN 사람을
바로/DAN 네가

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

거의/ADV 동났다.
거의/DAN 모든/DAN 공급품이

훨씬/ADV 예쁘다.
훨씬/DAN 전방/NNC+예/PAD+는/PAU

더/ADV 예쁘다.
더/DAN 이상/NNC 모른다.

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: 제/XPF+23/NNU
제 23: 제/DAN 23/NNU

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

제 단체: 제/DAN 단체/NNC

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

각급: 각/XPF+급/NNC

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

각 병과의: 각/DAN 병과/NNC+의/PCA

But ‘각종’ 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 ‘아’ 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 ‘아’ with ‘아서’: if it is grammatical and preserves the overall meaning, then ‘아’ has a similar function as ‘아서’, and, therefore, is an ECS.

먹어 보았다: 먹/VV+어/EAU 보/VX+았/EPF+다/EFN
같이 넣었다: 같/VV+아/ECS 넣/VV+었/EPF+다/EFN (= 같아서 넣었다)

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

먹게 되었다: 먹/VV+게/EAU 되/VX+었/EPF+다/EFN
먹게 했다: 먹/VV+게/ECS 하/VV+었/EPF+다/EFN
느리게 달렸다: 느리/VJ+게/ECS 달리/VV+었/EPF+다/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.

손이는 밥을 먹/VV+기는/ECS 학/VX+있/EPF+다/EFN.

날씨가 덥/VJ+기만/ECS 학/VX+다/EFN.

우리는 연구소의 시설을 사용/NNC+학/XSV+기로/ECS 뒤/VX+어/EAU 있다.

4.14 EFN or NNX

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

철수가 누구인지는 모르겠다: 누구/NPN+이/CO+ㄴ지/EFN+는/PAU

- However, in the following example, ‘ㄴ지’ 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 됐다: 가/VV+ㄴ/EAN 지/NNX+가/PCA

5 Confusing Examples

- ‘그럼’

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

- ‘대로’

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

명령대로: 명령/NNC+대로/PAD

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

말한 대로: 말한 대로/NNX

- ‘(이)라고’

If ‘(이)라고’ is attached to a noun which is a complement of a verb, it should be tagged as PAD as a whole.

완 닥/NPR+라고/PAD 부르 닥.

In some cases, ‘(이)라고’ 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+‘(이)라고’ is not a complete sentence but rather a sentence fragment; therefore there is no need to treat ‘(이)라고’ as a sentence final ending (EFN) and is tagged as PAD.

A: 그는 부산에서 삽니다.

B: 부산/NPR+이라고/PAD ?/SFN

- ‘말락’

In most cases, ‘말’ in ‘말락’ 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+지/EAU 막/VX+락/EFN
너 말/VV+고/ECS 네가 가락

- ‘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+하/XSF+금지도/ECS 모르/VX+ㄴ 닥/EFN.

- ‘밖’

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

너밖에 없다: 너/NPR+밖에/PAU 없다
창문 밖에: 창문/NNC 밖/NNC+에/PAD
그 밖에 무엇을 할까요: 그/DAN 밖/NNX+에/PAD

- ‘별’

‘별’ can be tagged as XSF or DAN.

부대별: 부대/NNC+별/XSF
별이 상한 소리: 별/DAN 이상한 소리

- ‘분’

‘이분, 그분, 저분’ are tagged as NPN. But ‘분’ can be a separate word, and in this case, it should be tagged as NNX.

급 하신 분: 급 하/VJ+시/EPF+ㄴ/EAN 분/NNX

- ‘본

If it means ‘original’, then it is tagged as XPF.

본 마음: 본/XPF+마음/NNC

본 바탕: 본/XPF+바탕/NNC

If it means ‘this’, it is tagged as DAN.

본 사건: 본/DAN 사건/NNC

- ‘상’

When ‘상’ has the meaning of ‘appearance’, it does not get a separate tag.

어머니상/NNC

아버지상/NNC

When ‘상’ means ‘above’ or ‘according to’, it should be tagged as NNC:

지도 상 예: 지도/NNC 상/NNC+예/PAD

- ‘소’ (place)

If ‘소’ occurs with the dependent noun ‘개’, it is tagged as XSF.

3 개소: 3/NNU 개/NNX+소/XSF

But when it occurs in ‘휴게소, 발전소, 사무소, 출장소, 훈련소’, it does not get a separate tag, and the words as a whole are tagged as NNC.

- ‘수’

‘수’ associated with a noun is a prefix. But if it occurs as a separate word, it is tagged as DAN.

수 십 개: 수/XPF+십/NNU 개/NNX

수 킬로미터: 수/DAN 킬로미터/NNX

- ‘셈’

‘셈’ is a common noun (NNC) when it means ‘calculation’. But it can also function as a dependent noun (NNX), in which case it does not mean ‘calculation.’

셈이 느리다: 셈/NNC+이/PCA 느리다
되는 셈: 되는 셈/NNX

- ‘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+이 없다.
공부하는 시간/NNC

- ‘요’

‘요’ is tagged as PAU if it attaches to a noun, but it is tagged as EFN if it attaches to a verb.:

나는 요: 나/NPN+는/PAU+요/PAU
갔어요: 가/VV+았/EPF+어요/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:

학교다: 학교/NNC+이/CO+다/EFN

In many contexts where ‘noun+락’ occurs, ‘이’ must be identified between the noun and ‘락’. Adnominal ending ‘락는’ presents one such case:

완다락는 물고기: 완다/NPR+이/CO+락는/EAN 물고기
철수가 완다락는 사실: 철수가 완다/NPR+이/CO+락는/EAN 사실

Contrast the above example with ‘락는’ with seemingly similar examples with ‘라고’ below. Copula ‘이’ is identified/recovered only when ‘라고’ attaches to a predicative noun in a clause, and not when it attaches to an argument noun. In ‘완다라고 부른다’, ‘라고’ is tagged as PAD, enabling us to treat ‘완다라고’ as the internal argument of the verb ‘부른다’. On the other hand, in ‘나는 철수가 완다라고 믿는다’, ‘철수가 완다라고’ forms a clause and therefore a predicate head is needed; copula ‘이’ is supplied for this reason.

완다라고 부른다: 완다/NPR+라고/PAD 부른다
 나는 철수가 완다라고 믿는다: 철수가 완다/NPR+이/CO+라고/ECS 믿는다

Moreover, ‘라고’ is tagged as ECS in the context where copula ‘이’ is recovered, while it is tagged as PAD if it is on an argument noun. Note that 이 in ‘이라고’ on an argument noun is not a copula, but is merely an epenthetic vowel. And so, ‘이라고’ in this case is tagged as PAD.

존슨이라고 부른다: 존슨/NPR+이라고/PAD 부른다
 나는 철수가 존슨이라고 믿는다: 철수가 존슨/NPR+이/CO+라고/ECS 믿는다

- ‘이니’

In most cases, ‘이니’ 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.

해방/NNC+이니/PAD, 통일/NNC+이니/PAD, 그 계 무슨 말/NNC+이/CO+니/EFN?

- ‘지’

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

빨개지다: 빨강/VJ+아/EAU 지/VX+다/EFN
 예뻐지다: 예쁘/VJ+어/EAU 지/VX+다/EFN

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.

버려지다: 버려지/VV+다/EFN
 나누어지다: 나누어지/VV+다/EFN

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.

공부되어지다: 공부/NNC+되/XSV+어/EAU 지/VX+다/EFN
 먹이어지다: 먹히/VV+어/EAU 지/VX+다/EFN

- ‘쪽’

‘동쪽, 서쪽, 남쪽, 동북쪽, ...’ are tagged as NNC as a whole. But ‘쪽’ can be a separate word and it is tagged as NNX in this case.

학교/NNC 쪽/NNX
선생님/NNC 쪽/NNX

- ‘측’

‘측’ is a suffix. But it can also be a separate word and it is tagged as NNX in this case.

주최/NNC+측/XSF
우티/NPN 측/NNX+의/PCA 의견/NNC

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

- ‘탁’

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

탁/XPF+부대/NNC
탁/NNC+의/PCA 모범/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.

밥을 먹/VV+어야/ECS 하/VX+ㄴ다/EFN
밥을 먹/VV+으려/ECS 하/VX+ㄴ다/EFN

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

밥을 먹/VV+기도/ECS 하/VX+ㄴ다/EFN
밥을 먹/VV+기만/ECS 하/VX+ㄴ다/EFN

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

손이는 철수가 춤을 추/VV+게/ECS 하/VV+였/EPF+다/EFN

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

손이는 노래/NNC+를/PCA 하/VV+였/EPF+다/EFN

If ‘하타’ is used as a light adjective, it is tagged as VJ.

손 이 가 쾌 왓 /NNC+은/PAU 하/VJ+타/EFN

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

손 이 는 철 수 가 바 보 /NNC+이/CO+라고/ECS 하/VV+있/EPF+타/EFN

- ‘한’

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

한/NNU 사람 만 갔 다 .

한/DAN 2000 미 터 정 도 됩 니 다 .

- ‘호’

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

1 호 : 1/NNU 호 /NNX