Vowel Sound Disambiguation for Intelligible Korean Speech Synthesis

Ho-Joon Lee
Computer Science Division
EECS department, KAIST
373-1 Guseong-dong, Yuseong-gu
Daejeon 305-701 South Korea
hojoon@nlp.kaist.ac.kr

Jong C. Park
Computer Science Division
EECS department, KAIST
373-1 Guseong-dong, Yuseong-gu
Daejeon 305-701 South Korea
park@nlp.kaist.ac.kr

Abstract

For speech synthesis systems that transform text materials into voice data, correctness and naturalness are the crucial measures of performance, the latter gaining more emphasis recently. In order to make synthesized voices natural, we must take into account pronunciation-related linguistic phenomena such as homograph, among others. The syntax certainly provides an important clue to disambiguating such homographs, but the relatively free word order in the Korean language makes it hard to utilize such information. In this paper, we describe a computational generation of contextually appropriate vowel lengths for the words in Korean by utilizing a higher level of linguistic information in a Combinatory Categorial Grammar framework. We consider parts-of-speech information, the possibility of conjunction with a suffix, case information, unconjugated adjectives, numerals, numerical adjectives with related nouns, and the relationship between a noun and its predicate as syntactic and semantic clues for vowel sound disambiguation. The results are expressed in Speech Synthesis Markup Language (SSML) for a target system neutral application. The proposed system with correctly predicted vowel sound can be used not only as an educational tool, but also as a plug-in for enhancing the intelligibility of a general purpose Text-to-Speech (TTS) system.

1. Introduction

Speech plays an important role as a communication method from birth to death, and it has also been regarded as the most user-friendly interface to machines thanks to the great development of the Information Technology such as the technologies for computers and mobile phones. For speech synthesis systems that transform text sources into voice data, correctness and naturalness are considered the crucial measures of performance. The naturalness enhances usability of speech synthesis systems with adding intelligibility to synthesized results. In order to make synthesized voices more natural, we must take into account pronunciation-related linguistic phenomena, such as homograph, among others.

While most homographs in English have different vowel sound such as ‘read [ri:d]’ versus ‘read [red]’ and ‘bass [beis]’ versus ‘bass [bæs]’, homographs in Korean share nearly the same pronunciation with one another, except for the lengths of their vowel sound. This has apparently caused many people even including television news announcers, to suffer from the confusion between short vowel sound and long vowel sound in homographs.

Disambiguation of short and long vowel sound for a homographic word in Korean represents not only the standard pronunciation, but also a mark of different meanings of that word. For instance, the word ‘밤(pam)’ has short vowel sound ‘[bam]’ to mean ‘night’, and long vowel sound ‘[ba:m]’ to mean ‘chestnut’. The following examples also show two different meaning of the word ‘눈(nwun)’ that means ‘eyes’ with short vowel sound ‘[nun]’, and ‘snow’ with long vowel sound ‘[nu:n]’, respectively.

(1) 철수가 눈이 아프대.
(chelswuka nwuni aphutay: Chelswu has pain in his eyes.)
(2) 오늘은 눈이 올 것 같아.
(onulun nwuni ol kes katha: It is likely to snow today.)

The phenomenon that involves short and long vowel sound of Korean homographic words is hard to explain with simple phonetic rules alone. When a word is broken into several lexical units for analysis of its reaction with one another, vowel sound tends to be characterized either short or long even with the same morpheme and similar lexical item sequence. For further details, the reader is referred to Lee and Park (2005). It is thus more reasonable to assume that short and long vowel sound of homographic words can be addressed by lexical characteristics rather than by simple phonetic rules.

The correct disambiguation of short and long vowel sound not only gives rise to a standard usage for the word pronunciation, but also gives a chance to distinguish different meanings of a word that can not be inferred simply from the written sentence. For example, we can not identify the intended meaning and vowel sound length with following sentence alone.

(3) 나는 밤이 좋아.
(nanun pami coha: I love night / chestnuts.)

But when the above sentence is pronounced correctly by a speaker, the intended meaning of the involved word can be identified with proper short or long vowel length as in '[bam]' for 'night' and '[ba:m]' for 'chestnut'. More examples are shown in Table 1.[1]

Table 1: Word senses that depend on vowel lengths

Word	Long vowel sound	Short vowel sound
밤(pam)	chestnut	night
눈(nwun)	snow	eyes
말(mal)	speech	horse
굴(kwul)	tunnel	oyster
배(pay)	twice or more	abdomen, ship, pear
모자(moca)	mother and son	hat
부자(pwuca)	rich person	father and son
부채(pwuchay)	debt	fan
성인(sengin)	saint	adult
새집(saycip)	birdhouse	new house
유리(ywuli)	profitable	glass
천직(chencik)	lowly occupation	vocation as calling

During the disambiguation of homographs in written form, the syntax certainly provides an important clue, but the relatively free word order in the Korean language makes it quite hard to utilize such information. Note that the Japanese language, whose word order is nearly as free as Korean, has certain words with the same sound but with different Kanji characters. In this case, their meanings can be distinguished by the use of intervening characters such as 'あ', 'い', 'う', 'え', 'お', or '-'. For example, '歓呼(かんこ)' and '観光(かんこう)' mean 'acclamation' and

[1] For two-character words, only the first syllable is pronounced with long vowel sound.

'sightseeing', respectively. The presence of these characters has the effect of lengthening the original vowel sound. While it would not be so difficult to distinguish the right meanings of such words when we read the Kanji characters, the listener may still be easily misled if we pronounce them inaccurately during conversation. When we build a speech synthesis system for languages like Japanese, the ambiguity of the meanings of this kind can be resolved straightforwardly since we can distinguish the meanings of words by referring to the semantics of each Kanji character with the underlying lexicon.

The fact that vowel sound can be distinguished through context without the need for correct pronunciation tends to dissuade people from using correct vowel sound, but in practice, synthesized information with correct short and long vowel sound can boost the chances for listeners to understand the conveyed information even in a noisy and uninformative environment. And as we have already shown above, the words with correct vowel sound may deliver important information that can not be inferred from the context alone, including the intention of the speaker. Therefore, correctly disambiguated vowel sound makes synthesized results more natural and intelligible.

Homographic words in Korean generally appear in nominal form with different vowel sound lengths that have multiple meanings. In this paper, we focus on the nominal form of homographs with different vowel sound lengths, and describe the computational steps for disambiguating such word senses in a Combinatory Categorial Grammar (CCG) framework.

The results of vowel sound disambiguation can be directly applied not only to a speech synthesis system for enhancing its naturalness and intelligibility, but also to an educational system for training correct pronunciation of each homographic word.

The rest of this paper is organized as follows. The related work on Word Sense Disambiguation (WDS) and Korean intonation generation are briefly discussed in Section 2. In Section 3, we discuss the method of disambiguating different vowel sound of Korean homographs using the lexical characteristics of each word in a CCG framework. The implemented system is explained in Section 4.

2. Related work

2.1. Word sense disambiguation

In the natural language processing field, Word Sense Disambiguation (WSD) problems that have traditionally been addressed by knowledge based or statistical methods are now also dealt with by a hybrid approach exploiting the strengths of both approaches (cf. Ide and Veronis, 1998). Knowledge based approaches use lexical resources such as dictionaries and thesauri for knowledge extraction, and also grammatical information such as syntax and semantics for word sense disambiguation, whereas statistical approaches use statistical information of related words extracted from corpora for machine learning methods. In computer science, WSD problems are generally addressed with a series of machine learning algorithms such as Bayesian probability, neural network, and memory-based learning (cf. Cho, 1995 and Shin et al., 2001).

Homographs are traditionally regarded as a special case of word sense disambiguation problem, and are usually addressed by corpus based statistical methods. However, without the means for incorporating deeper linguistic information such as syntax and semantics into the system, statistical approaches alone can not provide adequate explanation for the linguistic phenomena at hand. For instance, we may have to use the parts-of-speech information for the determination of the proper vowel lengths for many homographs. And even when two homographs have the same (mostly nominal) parts-of-speech information, such further information as case, unconjugated adjectives, numerals, numerical adjectives with related nouns, the possibility of conjunction with a suffix, and the syntactic relationship with the predicate provide additional clues. In this paper, we propose to

use all the possible syntactic clues for vowel sound disambiguation. For this purpose, we use feature categories to represent lexical semantics with structurally organized forms.

2.2. Korean intonation generation

Most work on Korean intonation generation are usually based on statistical methods, where Lee (2000) adopted the bootstrap aggregating and born again trees to predict proper intonation. The proposed theory takes into account four prosodic components: boundaries, loudness, length, and intonational tunes. In order to model these prosodic components, the theory consults the linguistic information that influences intonation of Korean sentences, and extracts relationships between sentences and their possible intonation. The resulting system is reported to show high accuracy in predicting prosodic intonational information statistically.

Computational approaches that are concerned with different vowel sound of homographs include Lee et al. (2004). In their work, cooccurrence information extracted from a corpus such as WordNet entries and NTT noun hierarchy is used to calculate the semantic similarity between words. In addition, pitch accents and pause information, as well as vowel sound, are expressed in Speech Synthesis Markup Language (SSML) for target system neutral application.

3. Vowel length disambiguation in a Combinatory Categorial Grammar framework

3.1. Combinatory Categorial Grammar

Combinatory Categorial Grammar (CCG) is a combinatory extension to the pure categorial grammar (CG) (cf. Steedman, 2000). CCGs provide a natural explanation for complex linguistic phenomena such as nonstandard constituent coordination and long distance dependency with a generalized notion of linguistic constituency and constituent-by-constituent combination of linguistic strings. In particular, CCGs provide an elegant way to integrate several steps of analysis such as syntactic, semantic, pragmatic, and discourse analyses in one derivation step. Table 2 shows the CCG derivation rules for Korean as proposed by Cho and Park (2000). Figure 1 represents simple ‘Subject (영희는: Yenghuy) - Object (빵을: bread) - Predicate (먹었어: ate)’ structured Korean sentence with syntactic and semantic analyses in a CCG framework.

Table 2: Proposed derivation rules for Korean

Derivation Rules	Derivation Name (Symbols)
X / Y Y → X	Forward Application (>)
Y X \ Y → X	Backward Application (<)
X / Y Y / Z → X / Z	Forward Composition (>B)
Y \ Z X \ Y → X \ Z	Backward Composition (<B)
X → T / (T \ X)	Forward Type Raising (>T)
X → T \ (T / X)	Backward Type Raising (<T)
X conj X → X	Coordination (<ϕ^n>)
X / Y Y \ Z → X \ Z	Forward Crossed Composition (>B_x)

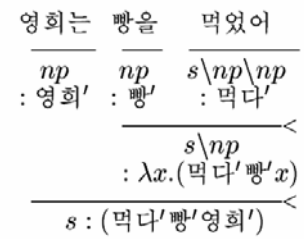

Figure 1: Syntactic and semantics analyses with a CCG

CCG categories for the Korean language are usually assigned on the eojeol (units separated by a blank space) level, or the most basic grammatical unit. But for more detailed information on and natural explanation for certain linguistic phenomena, it is more reasonable to assign CCG categories on the syllable or even morpheme level. In this paper, all CCG categories are assigned on the level of morpheme for the accurate assignment of syntactic and semantic information to each lexical item. For clarity of exposition and convenience, the derivations that are forthcoming contain chunked forms of morpheme-level categories. The extensibility of CCG categories to syllable and morpheme level is addressed by Lee and Park (2002) and Lee and Park (2003).

3.2. Vowel sound disambiguation for Korean homographs

3.2.1. Short and long vowel sound homographs with different syntactic structure

In many cases, homographs are easily distinguished of their short or long vowel sound by way of their parts-of-speech information. In example (4), '가장(kacang: most)' is used as an adverb with short vowel sound, whereas in example (5), '가장(kacang: the head of a family)' is used as a noun with long vowel sound.

(4) 가장 좋아하는 운동은 무엇입니까?
(kacang cohahanun wuntongun mwuesipnikka: What is your most favorite sports?)
(5) 그는 어려서부터 가장의 역할을 해왔다.
(kunun elyesepwuthe kacanguy yekhalul haywassta: He has played a role of the head of a family from when he was young.)

But the short or long vowel sound of homographic words is not always distinguishable with their parts-of-speech information alone. Examples (6) and (7) show the cases where homographic words '과장 (kwacang)' have the same parts-of-speech information.

(6) 그는 이번에 과장으로 발령받았다.
(kunun ipeney kwacangulo pallyengpatassta: He was promoted to the section chief.)
(7) 그는 이야기를 과장하여 말한다.
(kunun iyakilul kwacanghaye malhanta: He tends to overstate things.)

On the one hand, the homographic word '과장(kwacang)' in example (6) has the meaning of 'the section chief' with short vowel sound, but on the other hand, '과장(kwacang)' in example (7) is pronounced as long vowel sound with the meaning of 'overstate'. These two homographic words are all in nominal form, but the latter occurrence of '과장(kwacang)' in example (7) can be combined with a suitable suffix such as '-하-(-ha-)', whereas '과장(kwacang)' in example (6) can not. Thus, when a homographic word '과장(kwacang)' appears with a conjunctive suffix, the first syllable of this word will always be pronounced as long vowel sound. Many homographs including '소매(somay: sleeve / retail sale)', '장사(cangsa: business / powerful person)', '주의(cwuuy: principle / attention)' can be determined of their vowel sound by the possibility of conjunction with a suffix.

Homographs with different parts-of-speech information and the possibility of conjunction with a suffix are simple but effective syntactic clues for vowel sound disambiguation. In what follows, we look into more complicated syntactic restrictions on the cases where both short and long vowel

sound homographs are in nominal form with the same grammatical role such as subject, object, and complement[2].

3.2.2. Short and long vowel sound homographs both with in nominal form

Short and long vowel sound homographs that appear both in nominal form are hard to distinguish. Example (8) shows that the homograph '밤(pam)' can be translated into both short vowel sound '밤(pam: night)' and long vowel sound '밤(pam: chestnuts)'.

(8) 나는 밤을 싫어한다.
(nanun pamul silhehanta: I dislike night / chestnuts.)
(9) 나는 밤을 먹었다.[3]
(nanun pamul mekessta: I ate chestnuts.)

But in example (9), '밤(pam)' should be pronounced as long vowel sound by the characteristic of the verb '먹다(mekta: eat)' that takes '밤(pam)' as its object, which is a well known selectional restriction. The reason why '밤(pam)' in example (8) still has ambiguous meanings is that the verb '싫어하다(silehata: dislike)' can not restrict the properties of its object '밤(pam)' sufficiently. In examples (10), (11), and (12), each '벌(pel)' can be pronounced with its proper vowel length by the semantic restriction on each predicate.

(10) 벌이 날아다닌다. [벌:]
(peli nalataninta: A bee is flying.)
(11) 벌을 잡았다. [벌:]
(pelul capassta: I caught a bee.)
(12) 벌을 서다. [벌]
(pelul seta: I received punishment.)

The following three examples show the disambiguated vowel lengths with case information that each nominal homographic word has. We postulate that the case information syntactically influences its predicate phrase, and that the predicate phrase constraints the meaning of a nominal homograph.

(13) 호주가 가다. [호:주]
(hocwuka kata: The householder goes.)
(14) 호주를 가다. [호주]
(hocwulul kata: I go to Australia.)
(15) 호주에 가다. [호주]
(hocwuey kata: I go to Australia.)

The suitable vowel sound for nominal homographic words can be identified not only in consideration for the characteristics of predicate phrases like above examples, but also for unconjugated adjectives.

(16) 나는 달이 뜬 밤을 싫어한다. [밤]

[2] In this paper, we consider the essential adverbial phrase such as '학교에(hakkyoey: to school)' and '얼음으로(elumulo: into ice)' in "나는 학교에 갔다(nanun hakkyoey kassta: I went to school)" and in "물이 얼음으로 되었다(mwuli elumulo toyessta: Water was turned into ice)" as a complement.
[3] The poetic expressions like "I ate night." are not germane to the subject in this paper.

(nanun tali ttun pamul silhehanta: I dislike night with risen moon.)

Examples (16) and (8) are expressed with the same predicate '싫어한다(silhehanta: dislike)', but the presence of the unconjugated adjectives for the object '밤(pam)' in example (16) helps restrict the sense of the homographic noun to 'night' with short vowel sound. With a similar analysis, nominal homographic expressions in examples (18) and (19) are identified to have long and short vowel lengths, respectively, though homograph in example (17) is still ambiguous.

(17) 벌을 받았다.
(pelul patassta: I received a bee / punishment.)
(18) 죽은 벌을 받았다.[벌:]
(cwukun pelul patassta: I received a dead bee.)
(19) 엄중한 벌을 받았다. [벌]
(emcwunghan pelul patassta: I received a severe punishment.)

But in example (20), the homographic noun with the unconjugated adjective is hard to disambiguate for its vowel length, because the unconjugated adjective is not sufficient to restrict the semantics of the modified noun '벌(pel)'.

(20) 무시무시한 벌을 받았다.
(mwusimwusihan pelul patassta: I received scary bees / punishment.)

In addition, homographic nouns can be disambiguated by numerical adjectives, numerals, and numerical adjectives with related nouns. Homographic expressions in example (21) are disambiguated to have long vowel sound with the meaning of 'chestnut' by the numerals that follow, and homographs in example (22) are pronounced with short vowel sound with the meaning of 'night' by the numerical adjective.

(21) 밤 하나 (pam hana: one chestnut)
밤 둘 (pam twul: two chestnuts)
밤 셋 (pam seys: tree chestnuts)
(22) 한 밤 (han pam: one night)
두 밤 (twu pam: two nights)
세 밤 (sey pam: three nights)

The following examples describe the cases where each nominal homographic word is disambiguated by the associated numerical adjectives.

(23) 5 천만원의 배를 잃어버렸다. [배: / 배]
(5chenmanwenuy paylul ilhepelyessta: I lost twice of 50 million won. / I lost a ship which is 50 million won.)
(24) 5 천만원의 두 배를 잃어버렸다. [배: / 배]
(5chenmanwenuy twu paylul ilhepelyessta: I lost twice of 50 million won. / I lost two ships which are 50 million won.)
(25) 5 천만원의 배 하나를 잃어버렸다. [배]
(5chenmanwenuy pay hanalul ilhepelyessta: I lost one ship which is 50 million won.)
(26) 5 천만원의 배 한 척을 잃어버렸다. [배]
(5chenmanwenuy pay han chekul ilhepelyessta: I lost one ship which is 50 million won.)

The homographic word in example (23) can be analyzed to have both short and long vowel sound. Also in example (24), the homographic noun modified by the numerical adjective '두(twu)' can not be disambiguated. But by the numeral expression '하나(hana)' in example (25), the homographic noun can be pronounced with short vowel sound. Example (26)[4] shows that the presence of numerical dependent noun such as '척(chek)' gives a clue to disambiguate the vowel length of such a homograph, although a numerical adjective alone is usually not sufficient, as shown in example (24).

We have shown that a homograph can be disambiguated by taking into account the possibility of conjunction with a suffix, the syntactic relationship with a predicate phrase, case information, unconjugated adjectives, numerical adjectives, numerals, and numerical adjectives with numeral dependent nouns. In the section below, we propose to couch these analyses in a CCG framework.

3.3. Vowel sound disambiguation using Combinatory Categorial Grammar

Figures 2 and 3 show the CCG derivation steps for sentences in examples (4) and (5), respectively. As we have already mentioned in Section 3.2.1, different CCG categories are assigned to the homographic word '가장(kacang)' according to its POS information. In Figure 2, an adverbial CCG category '(S\NP)/(S\NP) is assigned to '가장(kacang)' to indicate that there is a verbal phrase '(S\NP)' on the right. Generally, a CCG category 'S\NP\NP' is assigned to transitive verbs (including a verb that takes complement as an argument), but in Figure 2, 'S\NP' is assigned to the verbal phrase '좋아하(cohaha)' to express the subject deletion condition. And for the predicate phrase '해왔다' in Figure 3 that has both subject and object as its arguments, the 'S\NP\NP' category is assigned.

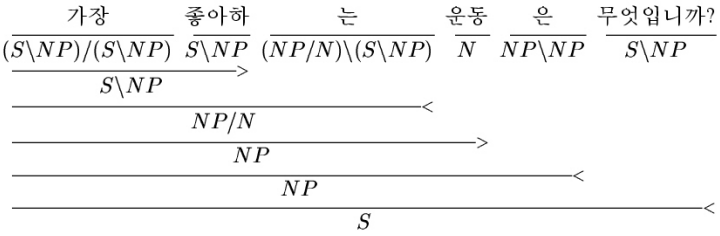

Figure 2: CCG derivation for example (4)

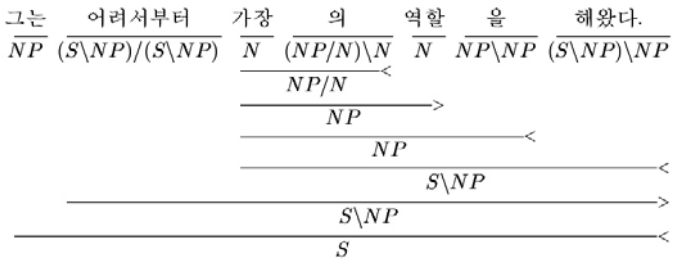

Figure 3: CCG derivation for example (5)

Figure 4 shows the derivation steps for example (6) where the homographic noun can not be combined with a suffix such as '-하-(-ha-)'. And Figure 5 illustrates a suffix combinable

[4] In this paper, we assume that '배 한 척(pay han chek: one ship)' is modified by the unconjugated adjective '5 천만원의(5chenmanwenuy: 5 million won)'. However, it is also a possible interpretation that '배(pay: ship)' is modified by the unconjugated adjective. Details for such instances of quantifier floating in Korean are discussed in Lee and Park (2003).

homographic noun with the CCG derivation. To express the possibility of conjunction with a suffix, we have incorporated '+suffix' and '-suffix' features into CCG categories. Therefore in Figure 5, '~하여(~haye)' can only be combined with a noun with the '+suffix' feature, whereas in Figure 4, '~으로(~ulo)' can be conjoined with every noun.

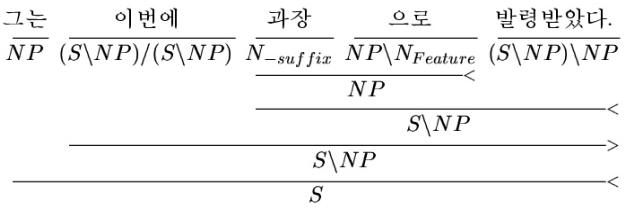

Figure 4: CCG derivation for example (6)

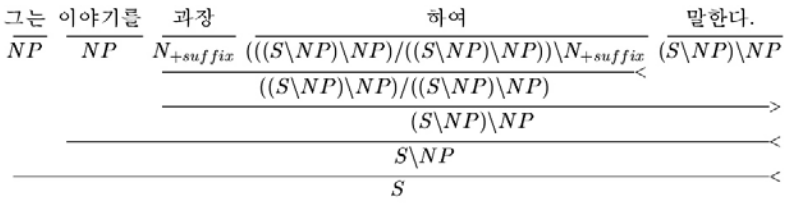

Figure 5: CCG derivation for example (7)

Figures 6, 7 and 8 illustrate CCG derivation steps for examples (13), (14) and (15), respectively, all of which can be disambiguated by the syntactic relationship between a predicate phrase and case information. In Figure 6, '가다(kata)' takes its nominal argument on the left, which is the subject, and plays the role of actor in the sentence. Because '가(ka)' assigns only subject case to its argument on the left, and does not change its role in the sentence, the argument (on the left) of '가(ka)' must have the '+actor' feature in its CCG category, to have long vowel sound.

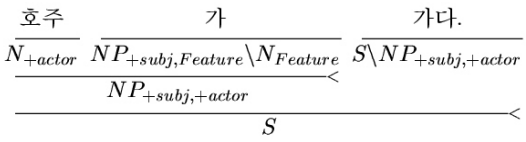

Figure 6: CCG derivation for example (13)

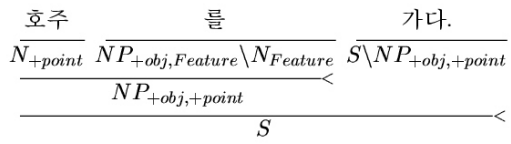

Figure 7: CCG derivation for example (14)

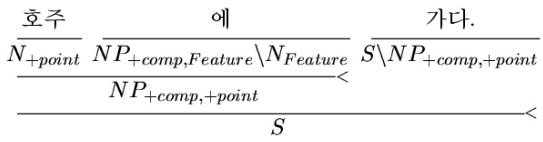

Figure 8: CCG derivation for example (15)

In examples (18) and (19), we have explained how unconjugated adjectives restrict their homographic nouns to have either short or long vowel sound. Their CCG derivation steps proposed in this paper are shown in Figures 9 and 10. In Figures 9 and 10, we represent the semantic structure of unconjugated adjectives and homographic nouns as features incorporated with syntactic categories, but it is also possible to express their semantic structures separately as illustrated in Figure 1.

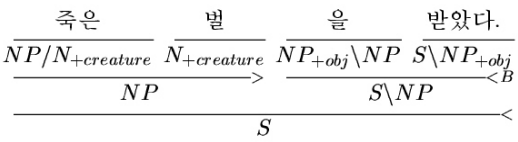

Figure 9: CCG derivation for example (18)

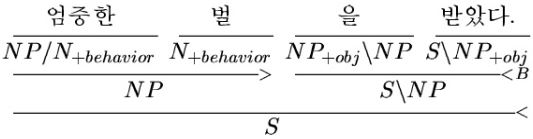

Figure 10: CCG derivation for example (19)

Figure 11 shows the derivation steps for the homographic noun '배(pay)' as modified by a numerical adjective '두(twu)', also shown in example (24), and Figure 12 illustrates the CCG derivation steps for example (26) where the homographic noun '배(pay)' is constrained by a numerical adjective with the resulting noun '한(han)' and '척(chek)' together. Though '두(twu)' in Figure 11 and '한(han)' in Figure 12 have the same syntactic CCG category, the homographic expression '배(pay)' in Figure 12 can be disambiguated by the numerical dependent noun '척(chek)'.

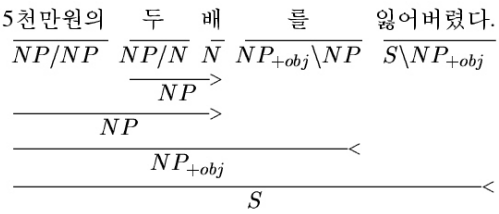

Figure 11: CCG derivation for example (24)

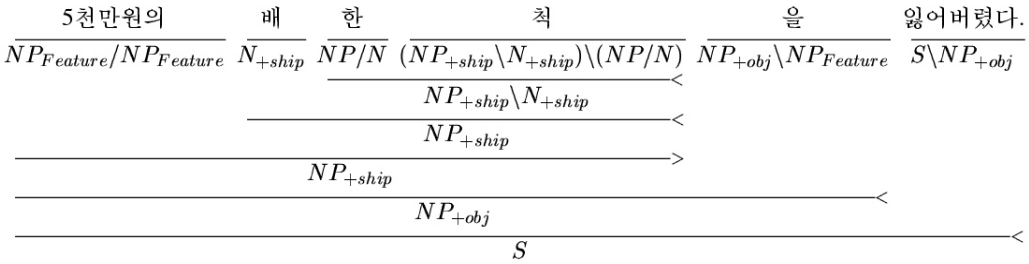

Figure 12: CCG derivation for example (26)

4. Discussion

In this paper, we have analyzed Korean homographs that can be pronounced with either short or long vowel sound, and described their disambiguating steps in a Combinatory Categorial Grammar

framework. We have investigated possible syntactic clues for vowel sound disambiguation, such as parts-of-speech information, the possibility of conjunction with a suffix, the syntactic relationship with a predicate phrase, case information, unconjugated adjectives, numerals, numerical adjective, and numerical adjectives with related nouns.

We have utilized morpheme-level CCG categories for detailed information of lexical items and natural explanation for vowel sound disambiguation. But for the clarity of exposition and convenience, chunked forms of morpheme-level categories have also been used.

The vowel sound disambiguated results are expressed in Speech Synthesis Markup Language (SSML) for target system neutral application. Figure 13 shows a piece of vowel sound disambiguated result as expressed in SSML. With the standard SSML expressions, properly generated documents can be applied to any other SSML-compliant TTS systems without additional conversion. Figure 14 shows the synthesized result generated by a ReakSpeak TTS system using a vowel sound disambiguated SSML document directly produced by the proposed system.

```
<?xml version="1.0" encoding="KS_C_5601-1987" ?>
- <speak xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="\\realvoicets\innovation\Preprocessors\data\doc\48MLsynthesis-
    core.xsd" version="1.0" xml:lang="ko-KR" ssft-dtype="ssft-text">
  - <p>
    - <sentence>
        그는 이야기를
        <prosody rate="slow">과</prosody>
        장하여 말한다.
      </sentence>
    - <sentence>
        <prosody rate="medium">호</prosody>
        주에 가다.
      </sentence>
    - <sentence>
        죽은
        <prosody rate="slow">벌</prosody>
        을 받았다.
      </sentence>
    - <sentence>
        <prosody rate="medium">배</prosody>
        한 척을 잃어버렸다.
      </sentence>
  </p>
</speak>
```

Figure 13: Generated SSML result with vowel sound disambiguated

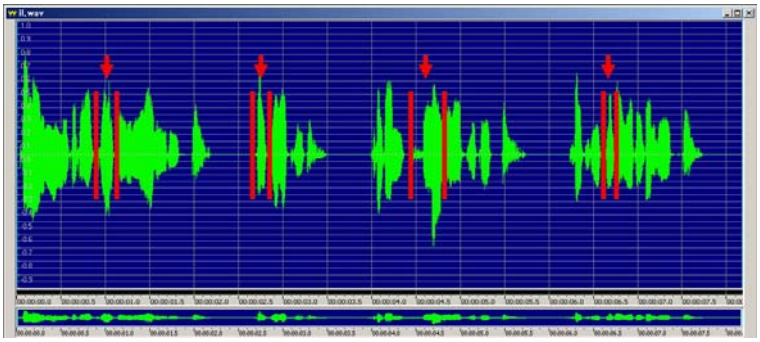

Figure 14: Speech synthesized result applying generated SSML document

The first and third arrows in Figure 14 indicate long vowel sound of '과(kwa)' in '그는 이야기를 과장하여 말한다(kunun iyakilul kwacanghaye malhanta: He tends to overstate things)' and '벌(pel)' in '죽은 벌을 받았다(cwukun pelul patassta: I got a dead bee)', respectively, and the second and fourth arrows point out short vowel sound of '호(ho)' in '호주에 가다(hocwuey kata: I go to Australia)' and '배(pay)' in '배 한 척을 잃어버렸다(pay han chekul ilhepelyessta: I lost one ship)', respectively.

The proposed speech synthesizer with correctly predicted vowel sound can be used not only as an educational tool, but also as an intelligibility enhancing plug-in for a general purpose Text-to-Speech (TTS) system. With properly generated vowel sound, the listener can easily figure out the word sense even in a contextually uninformative environment such as a noisy condition.

5. Acknowledgements

This research has been supported by Korea Ministry of Commerce, Industry and Energy through Brain Science Research Center.

6. References

Cho, Jeong-Mi. 1995. *Verb sense disambiguation using corpus and dictionary*. Ph.D. thesis, KAIST.

Ho-Joon Lee and Jong C. Park. 2002. Word Segmentation for Korean with Syllable-Level Combinatory Categorial Grammar. *Proceedings of the 14th National Conference on Korean Language Processing*, pp. 47-54.

Ho-Joon Lee and Jong C. Park. 2003. Morphological Analysis of Irregular Conjugation in Korean with micro Combinatory Categorial Grammar. *Proceedings of the Korea Information Science Society (KISS) Spring Conference*, pp. 531-533.

Ho-Joon Lee and Jong C. Park. 2005. Lexical Disambiguation for Intonation Synthesis: A CCG Approach. *Proceedings of the Korean Society for Language and Information (KSLI)*, pp. 103-118.

Hyung Joon Cho and Jong C. Park. 2000. Combinatory Categorial Grammar for the Syntactic, Semantic, and Discourse Analyses of Cordinate Constructions in Korean. *Journal of the Korea Information Science Society (KISS)*, 27(4), pp. 448-462.

Jin-Bok Lee and Jong C. Park. 2003. Analysis and Processing of Korean with Quantifier Floating, *Journal of Language and Information*, 7(1), pp. 1-22.

Lee Hwa Jin, Ho-Joon Lee, Jong C. Park. 2004. Constructing SSML Documents with Automatically Generated Intonation Information in a Combinatory Categorial Grammar Framework. *International Journal of Computer Processing of Oriental Languages*, 17(4), pp. 223-238.

Lee, Sang-Ho. 2000. *Tree-based modeling of prosody for Korean TTS systems.* Ph.D. thesis, KAIST.

Nancy Ide, Jean Veronis. 1998. Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. *Computational Linguistics*, 24(1) pp. 1-40.

Saim Shin, Juho Lee, Yongsoek Choi, Key-Sun Choi. 2001. Word Sense Disambiguation Using of Co-occurrence Information Vectors, *Proceedings of the Korea Information Science Society (KISS) Conference*, pp. 472-478.

Steedman, Mark. 2000. *The Syntactic Process.* The MIT Press.