Phonetic Transcription Rules and Quantitative Analysis of Phoneme Distribution in French

Hee-Sook Bae* · Young-Sun Yun** · Yung-Hwan Oh***

ABSTRACT

After establishing the rules for the phonetic transcription in French, quantitative analysis on the given text, *Waiting for Godot*, is performed. Analyzing the text by investigating the influence of phoneme distribution is very interesting in the phonostylistic point of view. Since the phonetic transcription rules are useful for its automation, the rules are carefully established in this paper. From the results of the phonetic transcription, we can investigate the distribution of individual phonemes and the different phoneme groups between dialogues and scenery indications for various characters.

Keywords: Transcription, Quantitative Linguistics, Phonostylistics

1. Introduction

To analyze the discourse, it is reasonable to explore the lexical distribution, first of all, because the importance of the word sense is widely known. However, the word is an entity with two aspects. The linguistic sign can be considered as the union of the sound and the sense, and these two elements tightly united.[1] The sense is not the unique factor that allows communication. For example, when a speaker utters or an author writes, a listener or reader produces not only the image of the signification of the words, but also the phonic image. The latter certainly plays a significant role for communication. In fact, the sounds are not all equally agreeable, equally tender, or equally quick; some phonemes are slow, others rapid; some are acute, others flattering; some are dark, others bright (P. Leon, 1993:12). In an aggressive and virile discourse, it can be imagined that the voiceless sounds, which contain acoustically acute and intense characteristics, will be employed more frequently than the voiced sounds that have tender and relaxed characteristics.

 * KORTERM (Korea Terminology Research Center for Language and Knowledge Engineering), KAIST
 ** School of Information Technology and Multimedia Engineering, Hannam University
 *** AI Center, Spoken Language Lab, KAIST
1) R. Jakobson (1984:39)

In this paper, we establish the rules for the phonetic transcription and then explore whether the phoneme distributions are strongly related to the thematic and dramatic structures.

Waiting for Godot of S. Beckett is selected as our corpus. In effect, S. Beckett is a playwright who has an interest in language plays, and this drama is his representative work. His work occasionally induces the audience to smile by the reversal of the presupposition:

    Estragon : Tu es sûr que c'était ce soir ? (...)
    Vladimir : Oui. Il a dit samedi. (...)
    Estragon : Quel samedi ? Et sommes-nous samedi ?
    Ne serait-on pas plutôt dimanche ? Ou lundi ? Ou vendredi ?

<div align="right">(Godot, p. 18)</div>

It is easy to find this kind of language play in his work. Furthermore, this play is still frequently presented in France. The play can thus be regarded as a significant key in understanding the contemporary language. We investigate the difference of phoneme distributions by applying the simple statistical method.

2. Criteria

Before starting the study of the phoneme distribution, phonetic transcription must be done. However, the transcription requires the rules be most rigorous because the data obtained by the unstable rules maybe not reliable. The important factor in quantitative linguistics is not choice of rules, but clear explanation to which rules are consistently chosen.

The transcription generates the inventory of the phonemes composing the text. The phoneme inventory helps to examine how phonemes are distributed in the whole text or in the various fragments, and to see their behavior. From this process, it can be observed how the phoneme structure is influenced by the lexical structure and vice versa. For example, the grammatical words will play a significant role in the study on the relation between lexical and phonic structures. In effect, it is well known that the frequency of certain phonemes is dependent directly on the grammatical words; the frequency of the phoneme /v/ is dependent on that of second person pronoun in plural forms, and the phoneme /ʒ/ is very dependent on the frequency of the word *je*. It is also thought that the negative words, such as *ne* and *plus*, surely influence the distribution of the phonemes /n/ and /p/. Therefore, it is necessary to select and establish appropriate rules

before the phonetic transcription.

The same word can be differently transcribed according to several factors; adopted style, position in the rhythmic group, consonant environment. Since these factors are variably changed, there are no rules that are satisfactory for all cases. Thus, at first, we roughly classify the cases into two types; one is related to the scenery indications, another to the dialogues. The dialogues will be transcribed according to the rules for the everyday phraseology, whereas the rules for the descriptive phraseology are applied to the phonetic transcription of the scenery indications.

French pronunciation Dictionary of Leon Warnant is referred to for the phonetic transcription. If some words are transcribed to multiple pronunciations with the mention "pfs", only the first pronunciation will be kept. For example, "crescendo" is transcribed into [kReʃɛndo] along with [kReʃɛ˜do] after the mention of "pfs". In this case, the latter will be neglected. However, the word is transcribed by multiple pronunciations according to the style, its pronunciations are all preserved. For example, "ennemi" is represented by two pronunciations: [enəmi] and [ɛnmi]. The first is the pronunciation used in a text with formal style, and the second is that of familiar style. Thus, in this paper, all cases of neutralization are neglected and the concept of archiphoneme is not used.

For the question of the distinction between [a] and [ɑ], in introduction of the dictionary, Leon Warnant said that many people have undoubtedly the two vowels at the present time; one is a little anterior, another a little posterior (1993: LXXVII). Therefore, he represented [ɑ] after [a] with braces and never used [a] in the primary transcription. However, we will use them without discrimination in this paper, like M. Hug (1979:13), who pointed out that the distinction between [a] and [ɑ] is often ambiguous and one seems only a stylistic variant of another in the statistical proportion.

Even if the phonetic transcription of each word follows the rules of L. Warnant and takes account of the word environment by considering different pronunciations, some problems still remain. First, the phonetic transcription is mainly made for representative form of a lemma, not for all forms of the lemma. Thus, our intervention is necessary to transcribe the variable forms by conjugation or by hand. In the second place, the behavior of [ə] is very unstable. The rules of [ə] vary with the style and the phoneme is positioned in parentheses showing optional sign in the dictionary. Consequently, we have to decide whether the phoneme is presented or not.

Depending on the characteristics of text to be transcribed, such as the dialogues or the scenery indications, different rules can be applied. The dialogues are transcribed like the everyday phraseology and the scenery indications are done as the descriptive phraseology. The rules concerning the descriptive style are based on the principles of *Traité de la prononciation du français* of P. Fouché and rules of the everyday style are started from those of the *Prononciation du français stadard* of P. Léon. The work of P.

Fouché is based on ordinary text style and the rules of P. Léon are primarily related to pronunciations of current conversation. In the third place, the rules of the liaison also vary with the styles and the environment. However, we cannot unfortunately find the general rules on the liaison. As we adopted two different rules for the dialogues and the scenery indications, two rules for liaison will be applied to two types of fragments.

Finally, we should determine the delimitation of the rhythmic group. This work is very important to solve the problems of [ə] and liaison because after the delimitation of the rhythmic group, it is possible to fix the rules which relate to [ə] and liaison. It is known that "a word can be pronounced differently according to its position in the rhythmic unit"(Fr. Wioland, 1983: III) and that the meaning of the word can be changed in the speech chain. However, the linguists suggest only general principles for dividing the sentences into rhythmic groups. This is the reason that the phonetic transcription starts with the delimitation of the rhythmic groups. After the delimitation, positions of words can be determined in the rhythmic group. The omission or conservation of the [ə] will be decided by its position. In the similar way, the decision relating to the liaison is based on the rhythmic group.

2.1 Rhythmic group

P. Leon (1979:65) defined the rhythmic group as a word group presenting an idea. It forms a semantic unit, which generally coincides with a grammatical boundary. This definition is very allusive. He did not propose precise rules. However, the delimitation of the rhythmic group induces some problems as it is shown in the introduction of L Warnant (1993: XCV): "the delimitation of the rhythmic group depends on the behavior of the speaker, the circumstances in which it is expressed, the intention and the temporary affect of the speaker". In this paper, the sentences are divided into the rhythmic groups following the principles of the French accent indicated in *Traité de la prononciation du français* of P. Fouché, while they do not exceed 7 or 8 syllables.

It is impossible to enumerate all principles that are used to delimit rhythmic groups. We will enumerate only the principles that are concerned in the cases that the units belong necessarily to the same rhythmic group. In other words, those units cannot be in boundary of the rhythmic group as follows:

    1) The determinant and the qualifying adjective cannot be separated from the nouns which are directly determined by the rhythmic group: *Estragon essaie d' enlever sa chaussure* (*Godot*, p.9), *sa* (his) and *chaussure* (shoe) must always be in the same rhythmic group;

    2) The personal pronouns subjects (*tu, il, ils*), complements (*me, te, se*), the nominal (*l'on*), the relative pronouns (*que, dont, ou*), the neutral

interrogative pronoun (*que*), and the prepositions cannot be in the final position of the rhythmic group, except if punctuation follows them, e.g., *Comment se fait-il que⋯* (*Godot*, p. 15);

3) The interrogative adjective and the verb *être*, or the interrogative and the substantive units cannot be separated: *quel est..., quelle idee!*;

4) The not-accentuated personal pronoun subject cannot be in final position in the following cases: (1) when it is immediately followed by a verb or a complementary pronoun, and (2) when it is reversed and followed by a monosyllabic without pause, e.g. *écrit-il bien?*, as if it is located between an auxiliary and the last participle, or between an semi-auxiliary and an infinitive: *as-tu terminé?; dois-tu parler?*;

5) The personal pronouns object (*le, la, les*) cannot be in the final position, e.g. *tu les connais* (*Godot*, p.69), except in the case of the inversion by the imperative: *soulevez-le!* (*Godot*, p.63);

6) The disjoined personal pronouns and the words, such as même, *seul, aussi,* form a rhythmic group. The plural form of disjoined personal pronouns and the words like *tous (toutes) or autres*, cannot be separated;

7) The semantic unit cannot be separated between its elements if possible: *de temps en temps* (*Godot*, p.19).

2.2 The [ə]

In the rhythmic group, it is noted that three positions of the [ə] can be available: the initial, the interior and the final. The [ə] behaves differently according to the position and the consonant environment. Sometimes it is never pronounced, or it must be done absolutely or optionally. These three possible realizations are considered as "obligatory", "optional" and "prohibited" as other linguists do. In the literary style, the [ə] will be pronounced in the obligatory and optional cases, while in the ordinary style it is restricted to only the obligatory case.

2.2.1 At the initial position of the rhythmic group

According to Fouché, the [ə], which is located at the initial position of rhythmic group, can be influenced by the preceding rhythmic group. For instance, the first [ə] is always preserved in the following cases; 1) the rhythmic group is not completely isolated from the precedent by a silence and 2) the last word of the preceding group ends in a pronounced consonant or a "consonant + e", "es" or "ent", If the last word of the preceding group ends in a vowel, the first [ə] is omitted. However, if the rhythmic group is divided from the precedent by a silence or punctuation, the following group is treated as if it was insulated or the initial part of a sentence. Nevertheless, we decide that the

presence or the absence of the first [ə] is determined only in the rhythmic group.

2.2.1.1 One syllable including the [ə]

If the [ə] is preceded by successive two consonants at the initial position of rhythmic group, it is pronounced obligatorily. However, if only one consonant locates ahead of the [ə], the conservation or omission of the [ə] depends on the style. In the formal style, the [ə] is always preserved at the initial position of the group, while in the case of the informal style, it is not preserved except in the following situations:

1) the preceding consonant is occlusive: [t], [d], [k], [p], [b] or [g], e.g. *de la tour Effel* [dəlatuRefɛl] (*Godot*, p.11);

2) the omission of the [ə] provokes the consonant gemination or the clash of two similar consonants by the conjunction; if the [ə] is omitted in *je change*, consonant [ʒ] is clashed with [ʃ]. To avoid this clash, the [ə] is preserved;

3) if the omission of the [ə] brings about the successive consonants at the initial position; *le grand chapeau*;

4) the following word of pronoun que begins with a consonant: *que se passe-t-il?*;

5) the [ə] is a distinctive value as in the words *dehors* and *dors*.

2.2.1.2 Two successive syllables including the [ə] together

If one syllable follows another syllable with the [ə] at the initial of rhythmic group, the first [ə] is preserved, while the second falls. This is the reason that the [ə] locating at the initial position of the rhythmic group in the formal style is determined to be always pronounced: *ne m(e) touche pas !* (*Godot*, p.81). However, in the informal style, the first [ə] is pronounced optionally, except some cases.

When the second [ə] precedes two consonants, a syllable starting with [nj] or [Rj] follows the second [ə], or the pronoun *je* is in front of the groups [sɥ], [Rɥ]: *que je suis*, the first and second [ə] are kept in both styles. However, if the first [ə] is preceded by a [l], [R], [m], [ɲ], [s] or [z] and the second by a occlusive: *ne me demande pas!*, the first is obligatorily marked, while the pronunciation of the second [ə] is optional according to the styles.

2.2.2 At the interior position of the rhythmic group

In this position, two cases can be considered roughly: 1) inside polysyllabic[2] or

---

2) In this paper, "polysyllabic word" includes the words that are polysyllabic with the pronounced [ə].

monosyllabic word that is proclitic; 2) at the final position of polysyllable.

2.2.2.1 Inside polysyllabic or monosyllabic word that is proclitic
2.2.2.1.1 One syllable including the [ə]
In the formal style, the [ə] is preserved if the [ə] is preceded by two or more pronounced consonants, but it is omitted if only one pronounced consonant precedes, even if this rule causes the succession of three consonants: *puis se précipitent* (Godot, p.28). However the [ə] is always preserved with following one pronounced consonant in the following cases:

    1) in front of [lj], [ʀj], [nj], [mj], [ʀɥ], [nɥ], [lɥ], [ʀw] or [lw]: *je lui dis*;
    2) inside polysyllabic word which begins with "consonant+[ə]" and which is
       preceded by a vowel: *en retire un navet* (*Godot*, p.26);
    3) in front of h aspired: *et se haussant*;
    4) to avoid the consonant gemination: *avant de dire*.

In the informal style, even if the [ə] is preceded by only one pronounced consonant, it is not conserved. However, it is preserved in front of h aspired or of consonant + [j], if it is a distinctive value as in the word *dehors*. If its omission is the cause of a series of several consonants, the [ə] is preserved to avoid the consonant successions. However, if three successive consonants and a semi-consonant follow one another, the [ə] is not pronounced: *qui le croit ?* (*Godot*, p.16).

If the [ə] is preceded by two or more pronounced consonants, it is preserved to avoid three successive consonants. However, even if the [ə] is preceded by two consonants, it can fall in the following cases:

    1) the second of the preceding successive consonants is [p], [t], [k], [b], [d],
       [g] or [v], and the following consonant is a [ʀ]: porterai [pɔʀtʀe] instead
       of [pɔʀtəʀe];
    2) the consonant group is easily articulated or very usual: *il sera, il fera*, etc.

2.2.2.1.2 Two successive syllables including the [ə] together
a) C+ [ə] + C + [ə]
In general, regardless of the style, the first [ə] falls and the second is preserved. However, the following cases deviate from the above rule:

    1) in the usual expressions: *je n(e), que n(e)*;
    2) if the first [ə] comes from a monosyllable and the next is from

polysyllables starting with a consonant + [ə], the first is kept and the second falls: *on ne s(e)rait*. However, in the formal style, both phonemes are preserved in the following cases; (1) the first [ə] is preceded by two pronounced consonants but the second following [z] is in front of [sɥ], or (2) the fall of the second induces the consonant gemination: *et que je suis...;*

b) C + [ə] + CC + [ə]

The first falls and the second is preserved: *dans le premier*. However, if the consonants are geminated or the clash of two identical consonants occurs, both [ə] are preserved.

c) CC + [ə] + C + [ə]

The first is preserved and the second falls except in the informal style. In the informal style, both [ə] are conserved:

1) if a word ending with "consonant + liquid + [ə]" is followed by a monosyllable le, ce, de: *tu peux prendre ce navet*;

2) if a word ending with "C + liquid + [ə]" is followed by polysyllables whose first syllable is consisted of "C + [ə]": *un example demande*.

d) CC + [ə] + CC + [ə]

Both [ə] are preserved.

2.2.2.1.3 At the end of a polysyllable in the rhythmic group

If the [ə] follows only one pronounced consonant and it precedes a word starting with one or two consonants, it is omitted: *la lun(e) pleine*. However, if it follows two pronounced consonants and it is followed by a word starting with a consonant, its behavior changes from the following cases; (1) if the following word is a monosyllable including the [ə], the first [ə] falls and the next is preserved: *de sort(e) que*, but (2) if the following word does not include the [ə] and starts with a consonant, the [ə] is omitted: *de court(e) jupe*. Sometimes, it is an hallucination to hear the [ə] pronounced after an occlusive consonant located at the final position of the word. However, it can be considered like a trace of explosion occurred at the two separated consonants. If the first [ə] is preceded by the word consisting of "constrictive + [k]" or made up with *-que*, it is also pronounced in the informal style: *quelques choses* (*Godot*, p.12), *quelques navets* (*Godot*, p.26).

However the pronunciation of the [ə] is obligatory in the following cases:

1) it follows the phoneme [z] and the following word starts with [s], [z] or [s]: *il pose ses chaussures…*;

2) it is preceded by "a consonant + [l] or [R]": *je rencontre personne*;

3) its omission provokes the geminated consonant: *il regarde dedans* (*Godot*, p.12).

The [ə] is pronounced if it follows two or three consonants and it precedes a word starting with two consonants: *le ventre plein* (*Godot*, p.25).

If a word that ends in "-e" precedes a word starting with a vowel, the [ə] is not pronounced inside the rhythmic group, regardless of the number of the preceding consonants: on lui offer un bouquet. However, it is conserved in the following cases:

1) in front of a word starting with *h* aspired;

2) in front of *oui*;

3) in front of the word *que* or *-que*: *puisque*.

2.2.3 At the final position of the rhythmic group

The final [ə] of the rhythmic group is always omitted even if it follows several pronounced consonants. However, it is pronounced in the accentuated words: *Prends-le!*.

2.3 The liaison

The liaison is the phenomenon that a final consonant of a word, normally dumb, is articulated in a rhythmic group if the following word starts with a vowel or semivowel. It is normally considered in the interior of the rhythmic group . The rules of the liaison can vary according to given style, the phonetic and syntactic environments. M. Hug (1995:229) presented the environmental factors that can influence the presence or absence of the liaison. These factors are as follows:

1) Characteristics of the phoneme that induces the liaison;

2) Position of a word inside the rhythmic group;

3) Number of syllables;

4) Identity of a word;

5) Grammatical feature of the preceding word (determinant, preposition);

6) Grammatical feature of the word that makes possible the liaison (verb, noun);

7) Integration degree of a word of foreign origin;

8) Syntactic relation between neighboring words.

Since the criteria for deciding the liaison are not simple, it is necessary to consider these factors to determine the liaison. In the previous section, two different standards are used for handling [ə] by characteristics of the text, which belongs to the dialogues or scenery indications. The standard of the liaison is based on the principles of M. Hug, which are established by a questionnaire and statistical tests. For the dialogues, the liaison is held only if it is obligatory, while the obligatory and the optional liaisons are considered for the scenery indications. Before beginning concretely processing the liaison, it is required to refer to the phenomena that can cause the liaison, i.e. modification of the phonemes.

2.3.1 The joint of the phonemes

If the phonemes adjoin, they are influenced with one another; the typical effects are lengthening, deafening, disnasalization, etc. According to P. Fouché (1959: 435), the change of timbre of a phoneme (*premier premier homme*: [pRəmje]-[pRəmjɛRɔm]) is a result of the liaison. However, P. Léon argued that this phenomenon results from the characteristic of the vowel [ɛ] in the unaccented position. In effect, this vowel is more opened than in the accentual position. By adopting the latter, the liaison induces only the disnasalization and the change of some consonants.

The disnasalization

In this case, it is not necessary to discriminate between the formal and informal styles because both styles treat this phenomenon in the same way. The nasal vowels [e] and [o] can become oral vowels or their nasality are preserved in the event of liaison. The nasal vowel [e] is disnasalized in the final position of the adjectival words ending with this vowel: un vain espoir. In the words *bien* and *rien*, the nasality of the [e] is also preserved: *rien* à *faire*. The vowel [o] always loses its nasality in the adjective *bon*: *un bon homme*. However, in the possessive adjectives *mon, ton, son*, they can preserve or lose their nasality: *mon ami* [monami] or [mõnami]. Therefore, from this viewpoint, we decided to keep the nasality of the vowel [o] in *mon ami*.

2.3.2 Change of the consonants

When the liaison takes place, its pronunciation corresponds to that of the final consonant in the preceding word. However, certain consonants take a different pronunciation from symbols of the orthography. Concretely, the phonetic symbol [d] is pronounced always [t] at the event of liaison: *le grand air* (*Godot*, p.33), the final "s" becomes always [z]: *ils étaient*, in the same way, the final "x" changed to [z].

2.3.3 The pronunciation of *donc*

The pronunciation of *donc* is very unstable. According to Petit Robert, this word is pronounced as [dŏk] in head of proposition or in front of vowel, while it is transcribed as [dŏ] elsewhere. By L Warnant, the pronunciation of this word is [dŏk] in the following cases; (1) if it leads a rhythmic group, it is in front of a vowel or consonant, or (2) if the speaker wants to emphasize on the word. However, in other cases, its pronunciation especially becomes [dŏ] after an imperative sentence or in an interrogative sentence.

M. Hug examined how this word was pronounced in the French by a statistical survey. After several tests, he concluded that the choice between the pronunciation [dŏ] and [dŏk] depended on the type of sentence (M. Hug, 1995: 1). In a declarative sentence, donc has almost an exclusively argumentative value, its pronunciation is normally [dŏk] (···); in an imperative sentence, [dŏ] was pronounced in front of a pause (···); finally in a interrogative sentences, it was always shown to be [dŏ] in front of a pause (···)." (M. Hug, 1993: 9). We apply the theory of Mr. Hug for the phonetic transcription of this word; in the declaratory sentence, this word will be marked [dŏk]: C' est donc a lui qu' il faut demander. (*Godot*, p.36), and in the imperative, interrogative, or exclamatory sentences, this word is pronounced as [dŏ] if it precedes a pause: *Vous attendiez donc ?* (*Godot*, p.31). Furthermore, if *donc* is employed in the last three types of sentence, it is pronounced as [dŏ] in front of a consonant: Mais *essaie donc de comprendre !* (*Godot*, p.22).

2.3.4 The obligatory liaison

The obligatory liaison takes place with regard to this style. The conditions of the obligatory liaison are considered in the following cases:

1) between a determinant (the article and other units alternating with the article) and a determined (the substantive, the substantival participle or the nominalized adjective): *les effets, les amis, les yeux, six heures, les Anglais,* etc. However, the numeral *vingt* and *cent* do not inevitably yield to this rule (Marc Hug, 1995:235).

2) between the epithet and the substantive or substantial word: *geste vers ses haillons* (*Godot*, p.14), *des trois autres* (*Godot*, p.15).

3) between the conjoint personal pronoun and the combined verbal form: *nous irons* (*Godot*, p.14).

4) between the personal pronoun and the adverbial pronoun, or between the adverbial pronoun and the verbal form: *Allons-nous-en !* (*Godot*, p.25).

5) between the verb and the postpositive and subjective pronoun in the interrogative sentences: Pourquoi criait-il ?, ayant allume sa pipe (Godot,

p.38).

6) between the one syllable forms of the verb *être* and its attribute of adjectival form or its complement: *la route est à tout le monde* (*Godot*, p.31), *Je suis heureux* (*Godot*, p.31). However, this rule is, in fact, less obligatory than the precedent items.

7) in the composed forms with the auxiliary *être: il est arrivé.*

8) between the compound words connected by a dash and inside the semantic units (or fixed expressions): *de plus en plus, tout à fait, tout à l'heure.*

9) after the adverb bien, the postpositive adjective or the past participle follow: *bien aimé, bien ententu.*

10) between the adverb *plus* and the following word: *plus agressif.*

11) between the adverb *tout* and the following word: *tout en noire* (*Godot*, p.50), *l' homme tout entier* (*Godot*, p.12), and between the subjective pronoun *tout* and the verb, if they are syntactically related to each other: *tout est comme ça* (*Godot*, p.12).

12) after the adverb *très*, the adverb, adjective or the past participle follow: *très agressif.*

13) the one syllable prepositions *dans, chez, en, sans* and the following word are always marked as the liaison: *dans un fosse, dans un sourire, sans indiscretion, sans importance, avec effort.* However, in front of a proper noun or the words *et* and *ou, sans* cannot be connected.

14) after the conjunction *quand* followed by a subjective pronoun *il(s), elle(s)* or *on: quand on les arrache* (*Godot*, p.21).

2.3.5 The optional liaison

The liaison optionally occurs in the formal style, while it does not in the informal style. In the following cases, the liaison is considered optional.

1) Between the plural substantive and the following words:
    a) the postpositive epithet or the adjectival participle: *les jambes écartées* (p.9);
    b) the conjugated verb: *les nuits arrivent;*
    c) prepositions only if the name and the preposition are syntactically dependent: *les yeux à demi fermés.*

2) Between the verb and the following word, if the latter is directly related to the preceding verb: *il se met à manger* (*Godot*, p.33). If the next word of the verb is its subject, e.g., the subject is inversed, a liaison does not occur. However, if the verb is *être* and the next is its attribute or

complement, the liaison is obligatory; in other cases, the verb *être* precedes an adverb, the liaison is optional: *c'est ici.*

3) Between the auxiliary or semi-auxiliary and the past participle: *je dois avoir quelques navets* (*Godot*, p.26). Nevertheless, if the auxiliary is *être*, the liaison is obligatory.

4) Between the adverb like *trop, toujours, plus, beaucoup, pas, bien, mieux, jamais, moins, souvent* and the following word: *toujours un peu plus* (*Godot*, p.52).

5) After some prepositions *après, avant, devant, depuis, pendant,* the words follows: *après un silence.* But there is no liaison when they precede a proper noun.

6) Between the conjunction *mais* and the following word: *Oui, mais en attendant* (*Godot*, p.21).

7) Between the adverb finished by "-ment" and the adjective or adjectival participle, the liaison is optional: *comment osez-vous ?* (*Godot*, p.46), *finalement avec succès* (*Godot*, p.51).

2.3.6 Prohibited liaison

The liaison mainly takes place in the same rhythmic group. The liaison after an accentuated word is thus not made. It can be said that if a punctuation mark follows a word or a pause is possible, the liaison is not made. However there are several cases where the liaison cannot occur, such as inside a rhythmic group.

1) In front of the interjection: *heu, ah, oh,* etc.

2) In front of the numeral *huit* and *onze*. However, inside a combined numeral, the liaison is often seen: *dix-huit, vingt-huit,* etc. Furthermore, the liaison can be made even in front of a consonant: *dix-neuf* [diznœf].

3) If the following word, which starts with *h*, is aspired.

4) Between the singular substantive
   a) and the postposition qualifying adjective (or the adjectival participle);
   b) and conjunctions *et* or *ou;*
   c) and the following verbal form: *le ton est bas.*

5) Between the qualifying adjective
   a) and the preposition or contracted article;
   b) and the following another qualifying adjective or past participle.
   In addition, the word *vingt* has a particular behavior in the combined numeral; sometimes, it is pronounced [vɛ̃t] even in the absence of the liaison: *vingt-six* [vɛ̃tsis], but it can be pronounced without [t] in

quatre-vingt-six. In the case of the liaison, the rule cannot explain the difference between the pronunciations of 21 and of 81: [vɛ̃teœ̃], but [katʁ əvɛ̃(e)œ̃].

6) Between the always accentuated indefinite pronoun, *l'un, les uns, les unes, quelqu'un, quelques-uns, quelques-unes, les autres, d'autres, quelques autres, chacun, certain, certains, plusieurs* and the following word: *l'un après l' autre.* However, the liaison is made in the set phrases: *l'un a l'autre, l'un et l'autre.*

7) Between the adverb such as *ailleurs, alors, assez, plutôt, longtemps, auparavant, autant, avant, debout, dedans, dehors, demain, dessous, dessus, devant, enfin, loin, maintenant, soudain, tard, tantôt* and the following word: aussitôt après.

8) After the preposition *selon: selon un déclenchement.*

9) Between the adverb like *cependant, pourtant, sinon* and the following word.

3. Phoneme Distribution and Stylistic and/or Thematic Characters

The text *Waiting for Godot* is transcribed into phonemes, and the frequency is calculated for each phoneme. Now, we can examine whether the distribution of phonemes influences the stylistic and/or thematic characters. How can it be examined whether the distribution of the phonemes influences the stylistic and/or thematic characters? This question can be solved by the statistical comparison of frequencies in individual phonemes and groups of phonemes. Since our corpus is composed of a theatrical play, it is adequate to start by the comparison between the dialogues and the scenery indications, and between the various roles.

3.1 Global phoneme distribution

The lexical study about *Waiting for Godot* (Hee-Sook Bae, 1990) presented that S. Beckett utilizes the same occurrences of the words for two fragments: the scenery indications and the dialogues. In this section, the phoneme frequencies are examined. From the result, the number of phonemes in the scenery indications is significantly less than that in the dialogues. The total number of phonemes in the dialogues is approximately twice that in the scenery indications. This quantitative contradiction between lexical and phonological units indicates that the word length composing the scenery indications is relatively short in this play.

Can this phenomenon be considered as the general information about the drama

language? Does it come from the difference between the narrative style and the dialogue style? Otherwise, is this characteristic due to writing style of S. Beckett? At this moment, it is simply supposed that this fact comes from an abundance of grammatical words (prepositions, articles, pronouns, etc.) more frequently used in the scenery indications than in the dialogues. In effect, the function words are generally smaller than the content words.

3.2 Dialogues and scenery indication

According to the observation of the phoneme distribution, the ranking of the phoneme frequency is similar in two fragments. However, scrupulous comparison shows the differences. Let us calculate the values to compare two averages,

$$Z = \frac{|X_a - X_b|}{\sqrt{\frac{S_a^2}{N_a - 1} + \frac{S_b^2}{N_b - 1}}}$$

In this equation, Xa and Xb show the averages to be compared, respectively; Sa^2 and Sb^2 are the corresponding variances; Na and Nb are the total number of blocks in which the variances are calculated. This formula makes it possible to see if the difference of the compared averages is statistically significant.

Table 1. The list of the phonemes and the corresponding z values

a	0.0982	ɑ̃	4.0839	g	6.6382	l	7.3947
ɛ	4.6598	ɛ̃	3.1558	k	2.2067	m	3.4736
e	2.2302	ɔ̃	0.0238	v	0.2967	n	8.3355
i	5.7587	œ̃	2.4186	f	1.8660	ɲ	0.8764
ɥ	0.9858	ə	1.0769	z	0.4888	w	6.6337
ø	4.4730	b	0.1756	s	3.3863	j	5.3523
œ	2.1714	p	2.8259	ʒ	5.0642	ɥ	0.7086
o	0.9224	d	0.3904	ʃ	4.3330		
ɔ	2.1452	t	1.1546	ʁ	5.2576		

Considering that the z value is statistically significant around 2.0, the investigation gives us interesting results; most cases are significant except for some vowels: /a/, /ɥ/, /o/, /ɔ̃/ and /ə/. The z value of /ə/ especially shows that the adopted principals don't remarkably influence the phoneme distributions. In addition, the most remarkable difference between the dialogues and the scenery indications are seen at the phoneme /n/;

8.3355. The phonemes /l/, /g/, /w/, /i/, /j/, /ʀ/, /ʒ/ are followed, and the phonemes having a value greater than 4.0 are /ø/, /ʃ/ and /a/. We can add three phonemes /s/, /m/ and /e/ to the remarkable difference list.

The distribution of the phoneme /n/ is directly related to the negative words: ne and non, which appear theoretically in the direct discourse and are related to the first person plural nous. However, the phonemes such as /i/, /l/, /w/, /ʒ/, /m/ are related to the personal pronouns and the definitive articles; the first two phonemes appear in the scenery indications, while the last three dominate the dialogues. The phonemes /ʃ/ and /s/ are related to the topic words: *chaussure* and *chapeau*.

It is natural that the scenery indications don't contain the first and second person pronouns, but have the third person pronouns. On the other hand, the frequency of the phonemes relating to the first and second person pronouns (i.e. /n/, /ʒ/, /m/, /w/) is considerable in the dialogues. All these phonemes are closely associated to the topic words, and play a significant role in seizing the topic of the play.

Let us examine the phonic characteristic of the two fragments by comparing the occurrence of voiced consonants with those of unvoiced consonants. According to Pearson's Chi-square test, there is no notable difference between the dialogues and scenery indications. However if we exclude the liquids from the investigation, the results become different; in the first degree of freedom, 50.008 and a corresponding probability is obviously lower than 0.001. The voiced consonants except the liquids are more abundant in the dialogues than in the scenery indications. By contrast, the scenery indications contain plenty of the unvoiced consonants.

Some explications can be extracted about this result by the lexical study, in which the most frequent words of the entire text are as following: *a, aller, arreter, attendre, avoir, bien, ca, ce, dans, de, dire, elle, en, Estragon, et, etre, faire, il, je, la, le, lever, Lucky, mais, mettre, mon, monsieur, ne, nous, on, oui, pas, plus, pour, pouvoir, Pozzo, que, qui, regarder, rien, savoir, si, silence, son, sur, temps, tu, un, venir, vers, Vladimir, voir, vouloir, vous, y*. Among these words, the words that are more abundant in the scenery indications than in the dialogues are: *a, arreter, de, en, Estragon, et, il, ils, le, lever, Lucky, mettre, Pozzo, regarder, se, silence, son, sur, temps, un, vers, Vladimir*. It can thus be considered that the abundance of unvoiced consonants is caused by the words *Estragon, Lucky, se, silence, son* in the scenery indications. The consonants constructing these words make the unvoiced consonants superior in the scenery indications. In addition, the words *chaussures* and *chapeau*, which are frequently used in the scenery indications and are considered as topic words of this play, are related to the frequency of /s/ and /ʃ/. This assumption can be confirmed by the examination of the consonants classified in three groups: the occlusive, the constrictive and the nasal groups.

Table 2. The groups of consonants

	Observed effectives			Calculated effectives		χ^2	
	Dialogues	S. Indication	Total	Dialogues	S. Indication	Dialogues	S. Indicaiton
occlusive	6847	3608	10455	6968.751	3486.249	2.127	4.252
constrictive	3462	2053	5515	3676.008	1838.992	12.459	24.905
nasal	2714	854	3568	2378.241	1189.759	47.402	94.754
Total	13023	6515	19538	6515	19538	185.899	

In the second degree of freedom, χ^2 is 185.899 and the corresponding probability is almost zero. The result strongly corresponds to our assumption; the occlusive and the constrictive are predominant in the scenery indications, whereas the nasal is very rich in the dialogue. The remarkable difference between the dialogues and the scenery indications appears as a contrast between the constrictive and the nasal.

The investigation of the distribution of the vowels, classified by their articulator position (front, intermediate and back vowels), can describe the phonic structure of this play in detail.

Table 3. Vowel groups

	Observed effectives			Calculated effectives		χ^2	
	Dialogues	S. Indicaiton	Total	Dialogues	S. Indication	Dialogues	S. Indicaiton
anteriors	8549	4534	13083	8607.330	4475.670	0.395	0.760
semi-anteriors	1270	553	1823	1199.355	623.645	4.161	8.002
posteriors	2191	927	3118	2051.338	1066.662	9.509	18.286
nasals	2577	1571	4148	2728.977	1419.023	8.464	16.277
total	14587	7585	22172	7585.000	22172.00	65.855	

In the third degree of freedom, χ^2 is 65.855 and the corresponding probability is lower than 0.001. Since the distribution of the vowel groups is uncertain, probability is extremely small. In table 3, the distribution of anterior consonants is nearly regular in these two fragments, and the nasal group is abundant in the scenery indications compared with the theoretical effective. However, the posterior group is rare in the scenery indications. If all vowels are examined without exception, it is difficult to draw the difference of sonority between the dialogues and the scenery indications.

For the nasal group, the abundance may come from the fact that the preposition *en* and the present participles finishing by -*ant* are frequent in the fragment. For a better measure of the sonority in the two fragments, we examined the distribution of the vowels by excluding the nasal group. In the second degree of freedom, χ^2 is 35.856 and

the probability is also lower than 0.001. If only the anterior and the posterior consonants are examined because their sonority is different, x^2 is 27.332 and the probability is lower than 0.001 in the first degree of freedom. This irregularity comes from the abundance of the posterior consonants in the dialogues. The posteriors are abundant in the dialogues.

The distribution of the classified vowels by the articulator position is similar to that of the classified consonants by the articulator mode. This result shows that the distribution of the classified vowels and consonants is very important in the dialogues and the scenery indications; the dialogues contain more the phonemes having tender, deep, thick and slow sonority. By contrast, the author uses the phonemes having aggressive, acute, sharp and quick sonority in the scenery indications more frequently. This result issues from the distribution of the phonemes /ʃ/ and /s/ in wide areas. The fact, which value increases in the test of the constrictive group, supports our assumption.

3.3 Dramatis personae

Through the comparison of the dialogues and the scenery indications, we knew that the phoneme frequency is influenced by the theme of discourse. However, this does not mean that the speaking style or the personality of characters also influences the phoneme frequency. To reveal the relation of them, the comparison of the roles can be applied. The comparisons are performed by the phoneme frequency and distribution.

From value which is calculated by the relative frequencies of the individual phonemes, we can know that the vowels /a/, /e/, /ɑ̃/ and the consonants /b/, /g/, /f/, /ʃ/, /ʁ/ and /l/ are most frequent in the role of Lucky. The most frequent phonemes in the role of Vladimir are the vowels /y/, /o/, /ɛ̃/, /œ̃/ and a consonant /t/. For Estragon, the phonemes /ɛ/, /ɔ̃/, /d/, /s/, /ʒ/, /w/, /j/ are frequent, and Pozzo employs the most frequently /œ/, /u/, /ə/, /p/, /v/, /z/, /m/. These results show that the frequent phonemes of Lucky are similar to those of the scenery indications. The phonemes often used by Vladimir are related to the words *nous* (we and us) and *tu* (you), those relating to Estragon with the word *je* (I), and those of Pozzo to the words *vous* (you) and *moi* (me).

In the role of Pozzo, *je* and *vous* are frequently used; in Vladimir, *tu* and *nous* are frequent. Besides, Estragon very often uses *je*. These phenomena are apparent in relation to the characters of each dramatis persona. It is well known that *vous* and *tu* are different in the stylistic level, and that Vladimir and Pozzo each employ the second person pronoun in different styles. However, the case of the first person pronoun is not simple. *Nous* is neither the simple plural of *je*, nor another expression of *je* in the semantic point of view. In a certain way, *nous* is opposed against *je*, and vice versa. It can be considered that Vladimir, who often employs *nous* and *tu*, has an extroversive nature, whereas the nature of Estragon using *je* is introspective. Let us compare the roles of characters by classifying the phonemes by sonority. The following table shows

well how the phoneme groups are spread over the four roles.

Table 4. Dramatis Personae (vowel groups)

	Vladimir	Estragon	Pozzo	Lucky
front	22.98	24.08	22.98	26.84
central	3.58	3.52	3.10	2.44
back	6.47	5.65	6.32	4.32

The anterior vowels having acute, aggressive and speedy sonority are predominant in the role of Lucky. If Lucky is excluded in the roles, Estragon frequently uses these vowels among the remaining roles. On the other hand, the role that frequently employs the posterior vowels of which sonority is deep, mild and slow is Vladimir and the next is Pozzo. From these results, it can be noted that sonority corresponding to each role is strongly related to their character; Vladimir is a role which is softer and less aggressive, and he speaks relatively slowly; Estragon speaks more quickly and aggressively.

The result relating to Pozzo and Lucky is a little different. In the play, the character of Pozzo is cruel and aggressive like a dictator, whereas Lucky is someone sacrificed. It can be considered from this result that Pozzo employs the words having an aggressive and cruel sonority, whereas Lucky uses the softer and slower sonority words. However, the distribution of the vowels classified by the sonority shows that Lucky speaks most aggressively, while Pozzo speaks more slowly and gently. This result reflects paradoxically enough the relation between Pozzo and Lucky, in other words between the directing and directed.

The nasal vowels appear frequently in the dialogues of Vladimir: 7.45, and the other roles follow in decreasing order: Pozzo (6.96), Estragon (6.79), Lucky (6.40). The result of the anterior vowels shows exactly inverse phenomenon. These results indicate that the most aggressive and acute role is Lucky and the next is Estragon, whereas Vladimir and Pozzo are the characters who relatively use softer and less aggressive discourses, and their discourses are relatively slow compared to Lucky or Estragon. Let us continue to examine the consonants, by gathering them based on sonority first of all into three groups (the occlusive, constrictive and nasal groups), and then two groups (the occlusive and constrictive groups) are once more divided into voiced and unvoiced.

Table 5. Dramatis Personae (consonants groups)

		Vladimir	Estragon	Pozzo	Lucky
occlusive	voiced	6.02	6.02	5.56	6.12
	unvoiced	13.06	13.04	12.70	13.60
constrictive	voiced	5.70	5.84	7.10	2.76
	unvoiced	7.61	7.80	7.61	7.92
	liquids	12.63	12.33	12.80	17.12
	nasals	7.16	7.42	7.66	6.96

The distribution of the consonants is also examined whether they have the same phenomenon. After the investigation, it is found that Lucky frequently uses the occlusive consonants. The fact, which the polarization of occlusive occurs in the role of Lucky, shows that the unvoiced consonants reflect well the character of this role. The occlusive consonants appear most rarely in the role of Pozzo. Furthermore, these phenomena correspond well with the results concerning the vowels; the constrictive voiced consonants are the most frequent in the role of Pozzo, whereas they are the least frequent in the role of Lucky; on the other hand, the constrictive unvoiced consonants are most often employed by Lucky, less often by Pozzo and Vladimir. In addition, the nasal consonants are a little more frequent in the role of Pozzo, whereas Lucky employs most rarely these consonants frequently using the liquids. The attentive observations on the individual phonemes give some analyses on the discourses of the roles: for /t/, it is the least frequently occurring in the discourse of Pozzo among the treated fragments, but /v/ is used most frequently. It is not very astonishing because Pozzo addresses *vous* for all the characters except his slave, whereas Estragon and Vladimir address *tu* for each other. For /ʒ/, Estragon and Pozzo employ it the most frequently, whereas Vladimir and Lucky seldom use it. This fact is similar to the result of /m/.

Now, we can think that the topic of this play appears clearer by comparing the dialogues and the scenery indications. As well, the style appears by comparison of the roles. The first demonstrates that the phonemes /s/, /ʃ/, /ʀ/, /m/, etc. are characteristically distributed and they are related to the topic words *chaussures, chapeau, enlever, mettre, redresser*, etc. The second reflects clearly that the sonority and the characteristics of the roles in its discourses, especially the comparison of the average frequency lead us to interpret as follows. The difference of averages of the same phoneme in both styles is very significant.

4. Conclusion

In this paper, we established the rules for the phonetic transcription in French. Since these rules will be very usable for future quantitative study on phonemes and for automation of phonetic transcription, they are carefully established in this paper. Depending on these rules, our text corpus *Waiting for Godot* was phonologically transcribed. In consequent, the phonemes inventory of two languages and their frequency are prepared. These data can contribute in two phases: 1) the phonemes inventory and their frequency can be considered as important research materials. In corpus-based research, we need statistic results extracted from various type of corpi to know the linguistic phenomena. 2) In favor of these data, we examined whether the phonemes distribution reflects the thematic and/or stylistic aspects of the text.

For the second phase, analyzing the text by investigating the influence of phoneme distribution is very interesting in the phonostylistic point of view. In fact, the result of the examination was very positive; the comparison of dialogues and scenery indications showed that scenery indication contains abnormally few phonemes related to the first and second person pronouns (/ʒ/, /m/, /n/) or to the words appearing mainly in the direct style (/n/ and /w/). In contrast, the phonemes related to the third person pronouns (/i/ or /l/), the negative adverbs (/n/), the preposition and the present participle (/a/) and the topic words *chaussure* and *chapeau*: /ʃ/ and /s/ are more abundant in the scenery indication. By examination of the sonority, the distribution of the consonants classified according to the articulation mode and the distribution of the vowels classified according to the articulation position coincide exactly; the voiced consonants and the posterior vowels having a tender, mild, deep and slow sonority are dominant in the dialogues, while the unvoiced consonants and the anterior vowels having clear, acute, aggressive and fast sonority are abundant in the scenery indications. Moreover, we confirmed similar results by comparing roles; Vladimir often uses the phonemes /t/ and /n/; Estragon, /ʒ/; Pozzo, /m/ and /v/. Phoneme distribution is closely related respectively to the personal pronouns: *tu(toi), nous, je, me(moi)* and *vous*. Then, we confirmed that this phenomenon reflects clearly the characteristics of each role.

Nevertheless, there are a few things to be considered. To draw a better conclusion about the distribution of the phonemes by which the author aimed at carrying out a concrete topic, more detailed statistical tests needs to be conducted. In addition, this study is performed on the data that are obtained by transcription of the written text. This study is also useful to determine the differences between the transcription by the rules and the real speech data. This will be our future work.

References

Beckett, S. 1952. *Waiting for Godot.* Paris. Minuit.

Bae, H.-S. 1997. *Phonetic, lexical and syntactic structures in two plays of J Tardieu.* Ph.D. Thesis. Strasbourg Univ.

Bae, H.-S. 1990. *The lexical distribution in Waiting for Godot of S. Beckett.* D.E.A. Thesis. Strasbourg.

Dell, F. 1992. *Les règles et les sons, Introduction à la phonologie générative.* Coll., Savoir: sciences, Paris, Herman, 297p.

Duchet, J.-L. 1986. *La phonologie.* Coll., Que sais-je ?, Paris, PUF, 127p.

Fouché, P. 1952. *Phonétique historique du français.* Paris, Klincksieck, 106p.

Fouché, P. *Traité de prononciation française.* Paris, Klincksieck.

Guiraud, P. 1954. *Les caractères statistiques du vocabulaire.* Paris, P.U.F.

Grammont, M. 1954. *Traité pratique de prononciation française.* Paris, Delagrave, 241p.

Hug, M. 1995. "Du caractère facultatif de certaines liaisons obligatoires." in *SCOLIA*, n3, 227-244.

Hug, M. 1993. "Validité et fiabilité des tests d'usage: l'exemple de la prononciation de ≪Donc≫." in *ARBAI*, Actes des 4e rencontres regionales de linguistique, Bale, 143-153.

Hug, M. 1995. "Le temoignage d'une base de données sur la double nature du mot donc." *Actes du Congrès de linguistique romane de Palerme.*

Hug, M. 1979. *Distribution of the French phonemes (Statistical studies).* Slatkine.

Hug, M. 1980. "Statistic Linguistics in France." in *COIL.*

Jacobson, R. 1976. *Six leçons sur le son et le sens.* Paris, Minuit, 121p.

Leon, P. & M. Leon. 1979. *Introduction à la phonétique corrective.* Coll., Le français dans le monde B.E.L.C. Paris, Hachette et Larousse, 98p.

Leon, P. 1987. *Prononciation du français standard.* Paris, Didier, 98p.

Warnant, L. 1987. *Dictionnaire de la prononciation française dans sa norme actuelle.* Duculot, Paris.

Wioland, F. 1987. *La rythmique du français parlé.* Publication de l'I.I.E.F., n° 7, Strasbourg.

▲ Hee-Sook Bae
   KAIST, KORTERM
   373-1 Guseong-dong Yuseong-gu Daejeon 305-70, Korea
   Tel: +82-42-869-8794 (O)   +82-42-861-2965 (D)
   E-mail: elle@world.kaist.ac.kr

▲ Young-Sun Yun
   School of Information Technology and Multimedia Engineering
   Hannam University,

133 Ojung-dong Daedeok-gu, Daejeon, 306-791, Korea,
Tel: +82-42-629-7569   Fax: +82-42-629-7843
E-mail: ysyun@mail.hannam.ac.kr

▲ Yung-Hwan Oh
KAIST, AI Center Spoken Language Lab
373-1 Guseong-dong Yuseong-gu Daejeon 305-70, Korea
Tel: +82-42-869-3516 (O)   +82-42-861-1043 (D)
E-mail: yhoh@bulsai.kaist.ac.kr