

TP 2 : CORÉEN PHONÉTIQUE

INTRODUCTION

Nous allons construire un logiciel qui traduit du coréen en alphabet phonétique. L'utilisateur va entrer le coréen dans une section de texte du logiciel et le logiciel va afficher le texte en prononciation phonétique. L'interface utilisateur est déjà construite, il vous reste à écrire le code qui effectue la traduction. Pour que la communication entre le code écrit et celui que vous allez écrire fonctionne, vous allez devoir écrire le code pour des méthodes dont le contrat est déjà rédigé.

DESCRIPTION

Vous allez trouver sur Moodle, un fichier contenant le code de départ pour la classe TexteCoreen. Cette classe contient la signature pour un constructeur et une méthode. Vous devez écrire le code pour ces méthodes. Vous allez devoir ajouter des méthodes et des classes au besoin. Le constructeur va recevoir une chaîne de caractères coréens. Vous utilisez cette chaîne pour construire une représentation interne des données. Ensuite, le code dans la méthode traduire va construire une chaîne de caractères représentant la prononciation de cette chaîne à partir de la représentation interne.

DESCRIPTION DES HANGULS

Un hangeul (caractère/syllabe coréen) est composé de 2 ou 3 Jamos (lettres). Le premier jamo est la consonne initiale (dénote c_i), le deuxième est une voyelle (dénote v_i) et le dernier (optionnel) est la consonne finale (dénote d_i). Chaque case de la chaîne de caractères en entrées contient un hangeul. Il est représenté par une valeur numérique entre 44032 et 55215. Il est possible de retrouver les jamos (c_i, v_i, d_i) qui constitue un hangeul (u_i) à l'aide des formules suivantes.

$$u'_i = u_i - 44032$$

$$d_i = u'_i \bmod 28$$

$$v_i = 1 + ((u'_i - d_i) \bmod 588) \div 28$$

$$c_i = 1 + u'_i \div 588$$

TRADUCTION

La traduction en prononciation peut être décomposée en 3 étapes :

1. Traduction simple des jamos en phonétique internationale.
2. Correction pour les liaisons entre deux hangeuls.
3. Traduction de l'alphabet phonétique en Unicode.

Les 3 tables suivantes contiennent la prononciation internationale pour chaque jamo.

(remarque : la consonne initiale 12 est silencieuse.)

Consonne initiale		
Numéro (c_i)	Jamos	IPA
1	ㄱ	k
2	ㄲ	k̚
3	ㄴ	n
4	ㄷ	t
5	ㄸ	t̚
6	ㄹ	r
7	ㅁ	m
8	ㅂ	p
9	ㅃ	p̚
10	ㅅ	s
11	ㅆ	s̚
12	ㅇ	
13	ㅈ	tɕ
14	ㅉ	tɕ̚
15	ㅊ	tɕʰ
16	ㅋ	kʰ
17	ㅌ	tʰ
18	ㅍ	pʰ
19	ㅎ	h

Voyelle		
Numéro (v_i)	Jamos	IPA
1	ㅏ	a
2	ㅓ	ɛ
3	ㅗ	ja
4	ㅜ	jɛ
5	ㅛ	ʌ
6	ㅝ	e
7	ㅟ	ja
8	ㅢ	je
9	ㅤ	o
10	ㅥ	wa
11	ㅦ	wɛ
12	ㅨ	ø
13	ㅩ	jo
14	ㅪ	u
15	ㅫ	wʌ
16	ㅬ	wɛ
17	ㅭ	y
18	ㅰ	ju
19	ㅡ	ʊ
20	ㅣ	ɰi
21	ㅣ	i

Consonne finale		
Numéro (d_i)	Jamos	IPA
0		
1	ㄱ	k̚
2	ㄲ	k̚
3	ㄴ	k
4	ㄷ	n
5	ㄸ	nte
6	ㄹ	nh
7	ㅁ	t̚
8	ㅂ	l
9	ㅃ	k
10	ㅅ	m
11	ㅆ	p
12	ㅈ	s
13	ㅉ	tʰ
14	ㅊ	pʰ
15	ㅋ	h
16	ㅌ	m
17	ㅍ	p̚
18	ㅑ	p̚
19	ㅓ	t̚
20	ㅕ	t̚
21	ㅇ	ŋ
22	ㅈ	t̚
23	ㅊ	t̚
24	ㅋ	k̚
25	ㅌ	t̚
26	ㅍ	p̚
27	ㅎ	t̚

Ensuite, les tables suivantes indiquent les remplacements à faire lorsque deux consonnes se suivent. Ce cas arrive entre la consonne finale d'un hangeul () et la consonne initiale du hangeul suivant. Voici les différents types d'entrées dans les tables :

- $p_1 + p_2$: p_1 est la nouvelle prononciation de d_i . p_2 est la nouvelle prononciation de c_i .
- p : p est la nouvelle prononciation de d_i . c_i n'est pas prononcé.
- Case vide : La prononciation de c_i et d_i est inchangé.

		Jamo finale du hangeul courant (d_i)								
		ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ
Jamo initiale du hangeul suivant (c_{i+1})	ㅇ	k	ㄱ+h	n	ㄷ		ㄹ	m	p	
	ㅇ	k'	ㄱ+h	n+h	ㄷ		ㄹ	m+h	p	
	ㄱ	k+k'		n+k	ㄷ+k		ㄹ+k	m+k	p+k	
	ㄴ	ㄴ		n+n			ㄹ+l	m+n	m+n	
	ㄷ	k+t		n+t	ㄷ+ㄷ		ㄹ+t	m+t	p+t	
	ㄹ	k+n		ㄴ+l			ㄹ+l	m+n	m+n	
	ㅁ	k+m		n+m	ㄷ+m		ㄹ+m	m+m	m+m	
	ㅂ	k+p							p+p	

		Jamo finale du hangeul courant (d_i)										
		ㄱ	ㄲ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ	ㅇ
Jamo initiale du hangeul suivant (c_{i+1})	ㅇ	s	s	ㄴ+h	ㄷ+te ^h		ㄷ+te ^h	k+h	ㄷ+te ^h	p+h	h	
	ㅇ	ㄷ		ㄴ+h	ㄷ+te ^h		ㄷ+te ^h	k'	ㄷ	p		
	ㄱ	ㄷ+k		ㄴ+k	ㄷ+k		ㄷ+k		ㄷ+k	p+k	h+k	
	ㄴ	ㄷ+n	n+ㄷ	ㄴ+n	ㄷ+n		ㄷ+n		ㄷ+n	p+n	h+n	
	ㄷ	ㄷ+ㄷ	ㄷ+ㄷ	ㄴ+t	ㄷ+ㄷ		ㄷ+ㄷ	k+t	ㄷ+ㄷ	p+t	h+ㄷ	
	ㄹ			n								ㄹ
	ㅁ	ㄷ+m		ㄴ+m	ㄷ+m		ㄷ+m	k+m	ㄷ+m	p+m	h+m	
	ㅂ	ㄷ+p										
	ㅅ		s+s			ㄷ+te ^h						
	ㅈ		ㄷ+te ^h									

		Jamo finale du hangeul courant (d_i)											
		ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ
Jamo initiale du hangeul suivant (c_{i+1})	ㅇ	k+s	n+te	ㄴ+h	ㄴ+k	ㄴ+m	ㄴ+p	ㄴ+s	ㄴ+t ^h	ㄴ+p ^h	ㄴ+h	ㄴ+s	
	ㄷ	k+ㄷ	n+te+t	n+ㄷ	k+t	m+t	p+t	ㄴ+ㄷ	ㄴ+t ^h	p ^h +t	ㄴ+ㄷ	ㄴ+ㄷ	

Finalement, les tables suivantes contiennent les séquences d'Unicode pour les caractères de l'alphabet international.

IPA	Unicode	IPA	Unicode	IPA	Unicode	IPA	Unicode
h	0x0068						
k	0x006B	k̥	0x006B, 0x0324	kʰ	0x006B, 0x02B0	k̚	0x006B, 0x031A
l	0x006C						
m	0x006D						
n	0x006E						
p	0x0070	p̥	0x0070, 0x0324	pʰ	0x0070, 0x02B0	p̚	0x0070, 0x031A
s	0x0073	s̥	0x0073, 0x0324				
t	0x0074	t̥	0x0074, 0x0324	tʰ	0x0074, 0x02B0	t̚	0x0074, 0x031A
ŋ	0x014B						
ɕ	0x0255			ɕʰ	0x0255, 0x02B0		
r	0x027E						

IPA	Unicode	IPA	Unicode	IPA	Unicode	IPA	Unicode
a	0x0061	o	0x006F	ø	0x00F8	ʌ	0x028C
e	0x0065	u	0x0075	ɛ	0x025B		
i	0x0069	w	0x0077	ɯ	0x026F		
j	0x006A	y	0x0079	ɥ	0x0270		

TESTS

Un fichier de testes unitaires (JUnit 5) va être disponible sur Moodle pour tester votre constructeur et la méthode de traduction. Pour réussir les tests, vous allez devoir implémenter des méthodes supplémentaires. Le contrat de ces méthodes avec leurs signatures est donné dans le fichier TexteCoreen.

EXEMPLE

Soit l'entrée suivante :

쑥누빔떡볶이

Ce qui va donner les 'codes point' suivants :

[50137, 45572, 48548, 45829, 46369]

Après l'application de la formule pour diviser en jamo :

[(11, 9, 1), (3, 14, 0), (8, 15, 8), (4, 2, 5), (4, 21, 13)]

Correspondance directe en alphabet international :

[(s, o, ḳ), (n, u, -), (p, wA, l), (t, ε, ntɛ), (t, i, tʰ)]

Remplacement pour 2 consonnes qui se suivent. Cela arrive entre le premier et le deuxième hangeul.

[(s, o, ɸ), (-, u, -), (p, wA, l), (t, ε, ntɛ), (t, i, tʰ)]

En Unicode :

[(0x0073, 0x0324, 0x006F, 0x014B), (-, 0x0075, -), (0x0070, 0x0077, 0x028C, 0x006C), (0x0074, 0x025B, 0x006E, 0x0074, 0x0255), (0x0074, 0x0069, 0x0074, 0x02B0)]

=

[0x0073, 0x0324, 0x006F, 0x014B, 0x0075, 0x0070, 0x0077, 0x028C, 0x006C, 0x0074, 0x025B, 0x006E, 0x0074, 0x0255, 0x0074, 0x0069, 0x0074, 0x02B0]

Devrait afficher :

şonupwAltentçtit^h

DIRECTIVES

1. Le TP est à faire seul ou en équipe de deux.
2. Code :
 - a. Pas de `goto`, continue.
 - b. Les `break` ne peuvent apparaitre que dans les `switch`.
 - c. Pour une méthode :
 - i. Un seul `return`.
 - ii. Additionnez le nombre de `if`, `for`, `while`, `switch` et `try`. Ce nombre ne doit pas dépasser 6.
3. Indentez votre code. Assurez-vous que l'indentation est faite avec des espaces.
4. Commentaires
 - Commentez l'entête de chaque classe et méthode.
 - Une ligne contient soit un commentaire, soit du code, pas les deux.
 - Utilisez des noms d'identificateur significatif.
 - Une ligne de commentaire ne devrait pas dépasser 80 caractères. Continuez sur la ligne suivante au besoin.
 - Nous utilisons Javadoc :

- La première ligne d'un commentaire doit contenir une description courte (1 phrase) de la méthode ou la classe.
 - Courte.
 - Complète.
 - Commencez la description avec un verbe.
 - Assurez-vous de ne pas simplement répéter le nom de la méthode, donnez plus d'information.
- Ensuite, au besoin, une description détaillée de la méthode ou classe va suivre.
 - Indépendant du code. Les commentaires d'entêtes décrivent ce que la méthode fait, ils ne décrivent pas comment c'est fait.
 - Si vous avez besoin de mentionner l'objet courant, utilisez le mot 'this'.
- Ensuite, avant de placer les **tags**, placez une ligne vide.
- Placez les **tag** @param, @return et @throws au besoin.
 - @param : décrivez les valeurs acceptées pour la méthode.
- Dans les commentaires, placer les noms de variable et autre ligne de code entre les tags `<code> ... </code>`.
- Écrivez les commentaires à la troisième personne, EN FRANÇAIS.

REMISE

Remettre le TP par l'entremise de Moodle. Placez vos fichiers '*.java' dans un dossier compressé de Windows, vous devez remettre l'archive. Le TP est à remettre avant le 29 mars 23 :55.

ÉVALUATION

- Fonctionnalité (9 pts) : les testes unitaires réussissent.
- Structure (2 pt) : veillez à utiliser correctement le mécanisme d'héritage et de méthode.
- Lisibilité (2 pts) : commentaire, indentation et noms d'identificateur significatif.