A Framework for an Ontology-Driven Multi-Lingual Transcription System with IPA Representation

Chakkrit Snae[1], Eri Hirata[2]
and Michael Brueckner[1]

Abstract

One of the important issues in computational linguistics is to design systems for speech recognition and machine transcription which can be used for various types of spoken data. In manual as well as in machine transcription in particular, names as such are of great importance when addressing people, locations, and objects. In order to communicate names between language communities with different writing systems we have to transcribe or romanise the names into the corresponding writing systems. Names tend to show a certain intrinsic grade of variation, and this is even more the case for their transliterated or transcribed forms. Correct transcription and transliteration of names is one of the major problems in cross-cultural communication. Available standard manual transcription systems are often used inconsistently, or simply not used at all. Many computer-based transcription systems use orthographic forms or pronunciation, rule based and statistical approaches.

In this paper the authors propose an automatic transcription model for multi-lingual transcription systems. This model uses four advanced methods or tools: (1) a syllable pronunciation and segmentation model with rule based multi-lingual pronunciation, (2) a rule based approach with IPA (International Phonetic Alphabet) representation that is converted to different writing systems, (3) an ontology of phonemes to capture the phonetic qualities of characters for different languages, and (4) a phonetic based name matching algorithm called Meta-Sound (a combination of Metaphone and Soundex algorithms) for constructing the thesaurus of transcriptions to get different name variations of a specific name dynamically. This algorithm is designed for different language specific naming conventions, and it helps to produce highly accurate matches. The model is not only used for romanisation but also for the transcription into other writing systems, e.g. the Thai writing system.

1. Introduction

Research into natural language human-computer interaction shows that it is being applied more and more frequently in information systems within major world language communities (Sef and Gams 2003). Keeping track of world trends in this field is, for smaller nations and languages, also important for their national identity as well as for their cooperation in the global electronic arena (Sef and Gams 2003). One of the greatest difficulties in this regard is communicating names, such as personal names (anthroponyms) or product names correctly. Names are an intrinsic part of

[1] Department of Computer Science, Naresuan University
e-mail: chakkrit.snae@gmail.com
[2] Department of English, University of Birmingham
e-mail: exh411@bham.ac.uk

culture as can be seen by the various ways in which names are used by different societies. For example:

- a typical US American anthroponym consists of three parts, the given name, the middle name, and the surname, with the surname as the most important part for sorting lists of names
- a Thai anthroponym comprises a given name and a surname only, but here the given name is the main part which is used to order lists of names.

Many cultural communities have elaborated various elements of names (see Figure 1) which they use or have used in different ways (Snae and Brueckner, 2006). Communicating names has not only to do with cultural differences but also with practical aspects, such as different writing systems. This problem leads to the process of transcribing or transliterating names from one writing system into another.

1. Initial and feudal names
▪ for male–Mr.
▪ for female
  -single female –Ms
  -married female – Mrs
▪ professional Occupation–Dr. , Prof. , ranking career
▪ honorary titles and feudal names for social status in society - Lord, Sir, Knight, Baron, etc.

2.First name, given name
▪ mixture of parent names
▪ monks
▪ naming system
▪ astrology
▪ family, e.g. great grandparents
▪ book/ dictionary of names
▪ using name that matches character e.g. actor/actress

3. Middle name
▪ grandparents or great grandparents
▪ parents

4. Surname
▪ inherited from family
▪ the king or Royal family
▪ parents surnames
▪ grandparents
▪ using name that matches character e.g. actor/actress

5. Nickname
▪ last syllable of given names, e.g. Chakkrit → Krit
▪ first syllable of given names, e.g. Robert → Rob
▪ called by family
▪ called by friend / local community

6. Artist name and pseudonyms
▪ first name only, like Sasha
▪ surname only, like Chernyim (Thai comedian)
▪ only nickname, like Prince
▪ only abbreviation, like –ky
▪ for artists, monks, etc.

Figure 1: Element of names.

In a strict sense, transcription is the process of matching the sounds of human speech to special written symbols using a set of exact rules, so that these sounds can be reproduced later. Transcription, as a mapping from sound to script, must be distinguished from transliteration, which creates a mapping from one script to another that is designed to match the original script as directly as possible.

Moreover, in order to transcribe a natural representation of sounds (phonemes) of a source language as their written representation (graphemes) in a target language there is a need for a general system of writing phonemes. The International Phonetic Alphabet (IPA) is widely used for representing sounds in different languages as can be seen in systems like IT-TELLS (Snae et. al, 2006). A different system is SAMPA, which is meant to be usable by standard keyboards. There are also local representations for this alphabet, for example in German (see Wells 1996). Of all these systems, the IPA is certainly one of the best candidates for phoneme-sized segmented-labeling (Gibbon 2006). Many computer-based transcription systems use orthographic forms or pronunciation, rule-based approaches and statistical approaches, but there is little research on transcription systems which could deal with multilingual data.

In this paper the authors propose an automatic transcription model for multilingual transcription systems, called SPION. The model uses four advanced methods and tools: (1) a Syllable Pronunciation and segmentation model with rule-based multilingual pronunciation, (2) a rule-based approach with an IPA (International Phonetic Alphabet) that is to be converted to different writing systems, (3) an Ontology of phonemes to capture the phonetic qualities of letters and characters (graphemes) for different languages and writing systems, and (4) a phonetics based Name matching algorithm called Meta-Sound as a combination of Metaphone and Soundex algorithms for constructing the concepts of transcriptions to get different name variations of a specific name dynamically. A system based on this model must include suitable module interfaces, so that it can be adapted easily to other languages.

This paper is organized as follows: In the following section, we identify the problems involved with the multilingual transcription, and we present a brief review of the literature. Section 3 outlines the system model and methodology leading to SPION, where we present the syllable segmentation and pronunciation mode, the grapheme-phoneme-grapheme conversion for source and target languages, the ontology of phonemes, and the name matching algorithm. This is followed by Section 4, in which we present the framework of SPION. In Section 5, the summary is given and the further work is outlined.

2. Literature Review

2.1 Issues in Multi-Lingual Transcription

Transcription is the phonetic or spelling representation of one language using the alphabet of another language. Practical transcription can be done into a non-alphabetic language. For example, in a Hong Kong Newspaper, George Bush's name is transliterated into two Chinese characters that sounds like “Bou-sū” (布殊) by using the characters that mean “cloth” and “special”. Japanese vocabularies are also frequently imported from other languages, primarily (but not exclusively) from English. The phonetic alphabet of Japanese specially used to render foreign words and loan words is called Katakana and requires romanization or transcription (Knight and Graehl, 1998). Many words from English and other Western European languages are borrowed by Japanese and are transcribed using Katakana, one of the Japanese syllabaries.

Names and especially anthroponyms can vary in spelling and order which causes problems in retrieving them. The issue of name variations becomes more problematic when dealing with names from other cultures because the sorts of variation that are permitted may not be the same as those permitted in other languages. Names vary between cultures which for a long time has been an obstacle for creating a single method for automatic name processing (Snae and Brueckner, 2006). For example, within each of the following cultures, such as Korean, Arabic, Hungarian, and Hispanic all the names given are permitted variants of the same name except the last one (Dematteis et al., 1998). For instance, it is difficult to specify which of the examples below are given names and surnames.

PARK DOE REE / PAG TO NI / TO NI PAG (Korean)
MOHAMMAD ALI ABD EL NADIR NUR EL DIN / IMHEMED ABDUNADEER NOOREDDINE / MHMD NUR ABD AL NADER (Arabic)
Eoetvoes Lorant / Roland Eoetvoes / Eoetvoes Roland (Hungarian)
ENRIQUE CESAR VELEZ ARGUETA ENRIQUE BELES, QUIQUE VELEZ A. E. C. ARGUETA (Hispanic)

Although the spelling variants of the name elements in the final name are acceptable, name order is crucial. In the Eastern naming system (e.g. Korean, Hungarian Japanese, Chinese,) the family name appears in the leftmost position and cannot move to the rightmost position. In the Western system, (e.g. English, German, French, Spanish, Italian, and American) the family name appears in the rightmost position. In Hispanic names, the family name is the penultimate element (VELEZ); the name furthest to the right of it (ARGUETA) may be dropped, but not the family name. ARGUETA, if it occurs alone, would therefore refer to another family (Dematteis et al., 1998).

Spelling variations are especially prominent in names from non-Roman writing cultures when such names have been transcribed to Roman characters (e.g. the romanised form of an Arabic name: NOOR EL DIN, NURELDIN, and NUREDDINE). In addition, English spelling with its many-to-many sound/letter correspondences contributes to the problem of romanization of non Western names. Dialectal differences, historical and phonetic spellings all make the English names somewhat unpredictable. The latter is even the case for English names. The following examples are given by Reaney and Wilson (1997).

COLWELL, COLWILL, COLLWELL
LEA, LEE, LEGH, LEIGH, LEY, LEYS, LAY, LAYE, LYE
THOMPSON, THOMSON, TOMSEN, TOMSON
WORCESTER, WORSTER, WOOSTER, WOSTEAR

Another issue for transcription is that same words can be transcribed differently under different systems. In many cultures, available standard transcription systems tend to be not used or used inconsistently. For example, the Mandarin Chinese name for the capital of the People's Republic of China is Beijing in the commonly-used contemporary system Hanyu Pinyin, and in the historically significant Wade Giles system, it is written Pei-Ching (Pinyin 2007).

In Arabic, for example, although there are transcription systems used by libraries and other official agencies, transcription tends to be far less predictable and highly inconsistent, even with a single individual. For example, an individual whose name is “ABD EL NADIR” and the name may be romanised on one occasion as ABDUL NADEER and on another as ABDUNNADIR. Both name representations are correct and accurate romanisation of the Arabic name.

For the Thai writing system there exists an official standard called Royal Thai General System of Transcription (1999) which is used for rendering Thai names into the Roman alphabet. It uses only straight letters for vowels, diphthongs and aspirated consonants, and does not indicate the length of a vowel and the five different tones. “Thai Romanization” (Aroonmanakun and Rivepiboon, 2004) is an automated tool which transcripts Thai names or terms into roman letters. Even in cultures in which transcription systems provide a reliable standard, personal interpretation, accommodation to the spelling of another culture or perceptual confusion can cause the spelling to deviate from the standard. Thus, for instance, the Thai name GOFF will vary with KOFF, because G and K are Romanisation alternatives from different transcriptions systems. An observed variant of GOUGH, however, are GOFF and GOFFE, representing the influence of spelling of English surnames by Reaney and Wilson (1997). The Thai Romanization tool represents KOFF as GOLF, where GOLF is actually a correct transliteration of two different Thai names: กอฟ and กอล์ฟ.

As for Japanese language, there are three different written forms: *Hiragana* and *Katakana* which are syllabic scripts (alphabet), and *Kanji* which is logographic characters borrowed from Chinese. *Kana* (*Hiragana* and *Katakana*) is often referred to as phonemic orthography, that is each kana symbol represents one phoneme (such as ア＝/a/) or a specific sequence of two phonemes (such as か＝ /k/ + /a/). The romanization of Japanese, which is normally called 'romaji', is the writing of Japanese language using the Latin alphabet. There are several different systems for romanization of Japanese. The main systems are Hepburn Romanization, Kunrei-shiki Romaji (ISO 3602) and Nihon-shiki Romaji (ISO 360 strict). The most commonly used Romanization system is the Hepburn system. The *Kana* romanization of Japanese (i.e. phonemic transcription) cannot represent every sound, and therefore it is not "pronounced as it is written and [not] written as it pronounced" (Halpern, 2002; 2006). Halpern (2006) suggests that the Kana syllabary cannot be said to be truly phonemic, since there is a possibility of irregular correspondence between grapheme and phoneme. For instance, the following problems are identified by Halpern (2002; 2006):

1. Vowel Devoicing: preceding vowel is devoiced
(e.g. commonly with /i/ and /u/ between voiceless consonants)
2. Nasalization of /g/: preceding /g/ is optionally nasalized, becoming [ɦ]
3. Nasal assimilation
4. Spirantization of affricates
5. Sequential voicing
6. Palatilization
7. Consonant Gemination
8. Vowel Glottalization

It is clear that there is a need for disambiguating sounds which are not represented by Romanization of *Kana*.

2.2 Related Work

Several systems have been proposed to solve the problem of automatic transcription for major and minor languages as well (Snae and Pongcharoen, 2007; Snae et al., 2006). As for Japanese and English, Qu et al. (2003) set up an automatic transliteration tool for information retrieval systems.

Syllables as suprasegmental concepts are shown in Kamholz (2005) who sets up an ontology of sounds and sound patterns. Syllable segmentation and pronunciation models for Thai language are shown by Snae and others (Snae et. al., 2006; Snae and Pongcharoen, 2007). Moreover, there are selection models for phonemes which need the establishment of a pronunciation model, as is set out by

Schultz et al. (2006). On the other hand, there are writing systems which have intrinsic difficulties in producing a phonemic representation, such as the Chinese writing system (Zheng et al., 2005). However, those writing systems are not dealt with in this paper.

Phonetics ontologies are becoming more and more powerful instruments for the exchange and networking of linguistic data, as can be seen, for example, in the GOLD project which tries to establish an ontology for descriptive linguistics (GOLD 2005). In addition, the EMELD project, Aristar presents a phonetics ontology which is already quite elaborated (EMELD, 2005).

3. System Model and Methodology of SPION

A concept and model is here proposed to overcome the problems of multi-lingual transcriptions listed above. Our system model has been designed using the following concepts:

- **S**yllable segmentation and **P**ronunciation model
- **I**PA
- **O**ntology of phonemes
- **N**ame matching algorithm

This model is referred to as SPION for cross-language transcription and is illustrated in some examples below.

3.1 Syllable Segmentation and Pronunciation Model

The model is to convert an input name into a syllable pronunciation based on the principle of language pronunciation rules, also called syllabification. This is an easy way to transcribe and merge source language syllables to target characters, in case no suitable syllable pronunciation is found in the database.

The following example shows a model of Thai syllable segmentation based on Thai pronunciation as described by Snae and others (Snae et al., 2006; Snae and Pongcharoen, 2007). Thai names/words are built from one or more syllables which may or may not have a meaning. Syllables are constructed from consonants and vowels. One syllable can have a meaning of its own, which in cases where there are two or more syllables to a name or word with a more complex meaning. Names are segmented into syllables using rule based syllable segmentation according to Figure 2, where C is an initial or final consonant, Vi is an initial vowel, Vm is a middle vowel, Vl is a final vowel and T is a tone marker.

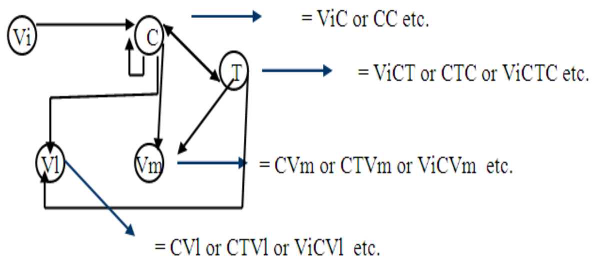

Figure 2: Thai syllables segmentation model.

Examples of the rule-based Thai syllables segmentation according to Figure 2 is: ใจ (Jai) = ViC, กล (Kon) = CC, เก่ง (Keng) = ViTC, ค่ำ (Khum) = CTVl, มา (Ma) = CVm.

The principles of Thai syllable segmentation model can be used for an English to Thai transcription which is a way of writing English words within the Thai writing system. While the letters of the English alphabet represent far more sounds than the Thai letters do, that is, its 26 characters represent 24 consonant and 20 vowel sounds, the Thai alphabet has far more letters than there are sounds: 44 characters represent 21 consonant sounds and 19 characters (including 3 consonant characters) represent 24 vowel sounds (including 6 diphthongs), 4 characters which are tone markers (่ ้ ๊ ๋) and a mark placed over the final consonant of a syllable to indicate that the letter is mute.

Thai syllable structures are different from those of English, so the phoneme sequences produced by the conversion rules have to be adjusted to comply with the Thai phonological system. Phoneme sequences which are not possible in Thai will be changed, deleted, or split into syllables in compliance with the Thai writing rules (Snae et al., 2006; Snae and Pongcharoen, 2007). As a consonant cannot stand alone in the Thai language, we consider rules for vowels only. The order follows basic rules, and examples can be found in Table 1.

Rule 1: Vowels which can come first or can be followed by a first consonant, e.g. Ek
Rule 2: Vowels which can follow a first consonant without a final consonant, e.g. Ka
Rule 3: Vowels which cannot take a final consonant, e.g. Tua
Rule 4: Vowels which need a final consonant, e.g. Kak

Rules of vowels	Classification of vowels	Romanisation	Examples
Rule 1	เ–, โ, แ– , ไ, ใ	e, o, ae, ai	Ek
Rule 2	(รร), (–า), (–ิ, –ี), (–ึ, –ื) (–ุ, –ู),	(an), (a), (i), (ue), (u)	Ka
Rule 3	–ะ, –ำ, เ–ะ, –ัว	a, am, e, ua	Tua, Tam
Rule 4	–ั, เ–็, ฤ, ฦ	a, e, i or e or oe, ue	Kek

Table 1: Examples of Classification of Thai letters.

The syllable pronunciation model which is applied uses syllable segmentation and the structure of Thai name patterns converts Thai names into syllabic pronunciation. In the following we describe how the algorithm of this model works: First we take an input name, and then delete characters which have the mark placed over the consonant, indicating that it is mute, e.g. จักรกฤษณ์ (input name) becomes จักรกฤษ. After that the remaining characters are segmented (using the syllable segmentation model) and arranged into word patterns. These word patterns are compared with the structure of the name patterns and then a syllable pronunciation is generated. As an example in Figure 3 the name วชิราพร can be segmented as follows: ว(C) , ชิ(C V middle), รา(C V middle), พ(C), ร(C) are structured as C V middle C V middle CC. Then this pattern is compared with the patterns in the structure of Thai names/words from left to right, in which only one pattern will match each word and then the correct syllable pronunciation will be generated (i.e. each matched word will be deleted from the pattern and the remaining characters will be compared). If it does not match any pattern then the first character will be deleted from the pattern and then checked for a match again, a process which is repeated until a match is found. After this, all the matches are brought to generate syllable pronunciation using "-" to separate each syllable.

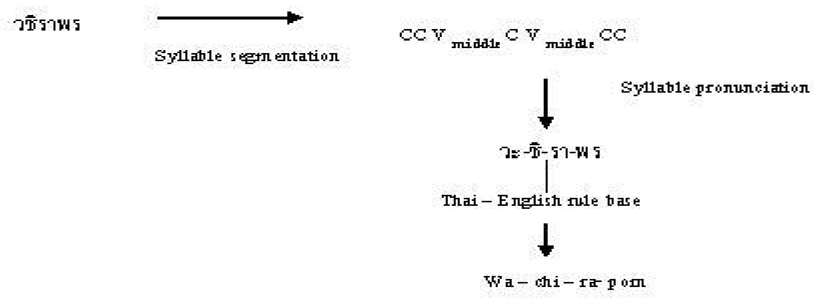

Figure 3: Syllable pronunciation using syllable segmentation.

We can use this concept for other languages by changing the rule based transcription and pronunciation based on the rules and customs of each language.

3.2 IPA Representations

Nowadays, in written correspondence, the localized representation of a foreign name is prioritized to the IPA representation. The same is true for many dictionaries in which intuitive representations are given not only for names but also for words. The IPA representation is useful for matching a phoneme and converting a phoneme into different writing systems and language communities. As an example, the allographs hat – hut – หัท have the same sound quality in three different languages, namely, German, English and Thai. The IPA representation is [h a t], a voiceless glottal fricative with an open central vowel and ending with a coronal alveolar plosive.

How can the set of phonemes be organized in such a way, that we can find easily which graphemes represent them in a given language and writing system? One way is to use a database or thesaurus of the phonemes in the IPA chart and match the different localized letters or characters. Keeping in mind that we want to use an automatic converting system to generate the intuitive transcription, we suggest using an ontology of phonemes which shows not only the corresponding sounds but also the graphemes representing the sounds in various writing systems.

In this paper, some examples are shown for working out the automatic cross-language transcription model. The examples refer to English, Japanese, Thai, and German. As the system deals with phonemes as the sound representations of written syllables or words, we need a system to represent phonemes unambiguously.

In the following we present some IPA examples for English (Table 2) German (Table 3 and 4) and Japanese and Thai (Figure 4 and Table 5), to set out the phonemes which are present in these languages.

IPA: English Consonants	
IPA	Examples
p	pen, spin, tip
b	but, web
t	two, sting, bet
d	do, odd
tʃ	chair, nature, teach
dʒ	gin, joy, edge
k	cat, kill, skin, queen, thick
ɡ	go, get, beg
f	fool, enough, leaf
v	voice, have
θ	thing, teeth
ð	this, breathe, father
s	see, city, pass
z	zoo, rose
ʃ	she, sure, emotion, leash
ʒ	pleasure, beige
h	ham
m	man, ham
n	no, tin
ŋ	singer, ring
l	left, bell
ɹ	run, very[1]
w	we
j	yes
ʍ	what[2]

IPA: English Vowels			
IPA			Examples
RP	GA	AuE	
ɑː	ɑ	aː	father
ɪ	ɪ	ɪ	sit
ɪ	i	i	city
iː	i	iː	see
ɛ	ɛ	e	bed[3]
ɜː	ɝ	ɜː	bird
æ	æ	æ	lad, cat, ran[4][5]
ɑː	ɑɹ	aː	arm
ʌ	ʌ	a	run, enough
ɒ	ɑ	ɔ	not, wasp
ɔː	ɔ	oː	law, caught[6]
ʊ	ʊ	ʊ	put, wood
uː	u	ʉː	soon, through
ə	ə	ə	about
ə	ɚ	ə	winner

IPA: English Diphthongs			
IPA			Examples
RP	GA	AuE	
eɪ	eɪ	æɪ	day, pain
aɪ	aɪ	ɑe	my, wise
ɔɪ	ɔɪ	oɪ	boy
əʊ	oʊ	əʉ	no, tow
aʊ	aʊ	æɔ	now
ɪə	ɪɹ	ɪə	near, here
ɛə	ɛɹ	eː	hair, there[7]
ʊə	ʊɹ	ʊə	tour
juː	ju	jʉː	pupil

Table 2: IPA examples for English variants[3] (Wikipedia, 2007).

Phonem	Beschreibung	Laut
/a/	kurzer, offener vorderer bis hinterer ungerundeter Vokal wie in Kamm.	[a]
/aː/	(langer,) offener vorderer bis hinterer ungerundeter Vokal wie in kam bzw. Kamin.	[a(ː)]
/ɛ/	kurzer, halboffener vorderer ungerundeter Vokal oder Schwa wie in Stelle bzw. bitte	[ɛ]/[ə]
/ɛː/	langer, halboffener vorderer ungerundeter Vokal wie in Käse.	[ɛː]
/eː/	(langer,) halbgeschlossener vorderer ungerundeter Vokal wie in stehlen bzw. Genom	[e(ː)]
/ɪ/	kurzer, fast geschlossener fast vorderer ungerundeter Vokal wie in Mitte.	[ɪ]
/iː/	(langer,) geschlossener vorderer ungerundeter Vokal wie in Miete, vital.	[i(ː)]
/ɔ/	kurzer, halboffener hinterer gerundeter Vokal wie in offen	[ɔ]
/oː/	(langer,) halbgeschlossener hinterer gerundeter Vokal wie in Ofen bzw. Roman	[o(ː)]
/œ/	kurzer, halboffener vorderer gerundeter Vokal wie in Hölle.	[œ]
/øː/	(langer,) halbgeschlossener vorderer gerundeter Vokal wie in Höhle bzw. Ödem	[ø(ː)]
/ʊ/	kurzer, fast geschlossener fast hinterer gerundeter Vokal wie in Mutter.	[ʊ]
/uː/	(langer,) geschlossener hinterer gerundeter Vokal wie in Mut bzw. Rubin.	[u(ː)]
/ʏ/	kurzer, fast geschlossener fast vorderer gerundeter Vokal wie in müssen.	[ʏ]
/yː/	(langer,) geschlossener vorderer gerundeter Vokal wie in müßig bzw. Physik.	[y(ː)]

Table 3: German vowels and IPA expressions (Wikipedia, 2007a).

3 RP received pronunciation, AuE Australian English, GA General Australian

Laut	Beschreibung	Beispiel
ʔ	Glottisschlag (Knacklaut) – Oft wird dieser Laut nicht als Phonem der deutschen Sprache beschrieben, sondern als morphologisches Grenzmarkierungsphänomen. In den südlichen Varietäten tritt dieser Laut nicht auf.	[bəˈʔaxtən] (nördliche Varietäten)
b	stimmhafter bilabialer Plosiv – Da dieser Laut in den südlichen Varietäten stimmlos ist ([b̥]), wird er oft als Lenis bezeichnet und nicht als stimmhaft.	Biene [ˈbiːnə, b̥iːnə], aber [ˈaːbɐ, ˈaːb̥ɐ]
ç	stimmloser palataler Frikativ (Ich-Laut) – Dieser Laut bildet zusammen mit [x] ein komplementäres Allophon-Paar. Er tritt nach vorderen Vokalen sowie nach Konsonanten auf. Im Diminutiv-Suffix [çən] tritt ausschließlich dieser Laut auf. Mit Ausnahme dieses Suffix' tritt [ç] in südlichen Varietäten im Silbenanlaut nicht auf, während es in anderen Varietäten oft im Silbenanlaut anzutreffen ist. In nicht-südlichen Varietäten ist [ç] ein übliches Allophon von /g/ im Silbenauslaut (nach vorderen Vokalen oder nach Konsonanten); die gemäßigte Standardlautung verlangt diese Spirantisierung nur in der Endung /ɪg/.	ich [ɪç], Furcht [fʊrçt], Frauchen [fʁaʊçən], nicht-südliche Varietäten: China [ˈçiːna], dreißig [ˈdʁaɪsɪç]
d	stimmhafter alveolarer Plosiv – Da dieser Laut in den südlichen Varietäten stimmlos ist ([d̥]), wird er oft als Lenis bezeichnet und nicht als stimmhaft.	dann [dan, d̥an], Laden [ˈlaːdən, laːd̥ən]
d͡ʒ	stimmhafte postalveolare Affrikate – Dieser Laut tritt nur in Fremdwörtern auf. In den südlichen Varietäten, die keine stimmhaften Plosive aufweisen, fällt er mit [t͡ʃ] zusammen.	Dschungel [ˈd͡ʒʊŋəl]
f	stimmloser labiodentaler Frikativ	Vogel [ˈfoːgəl], Hafen [ˈhaːfən]
g	stimmhafter velarer Plosiv – Da dieser Laut in den südlichen Varietäten stimmlos ist ([g̊]), wird er oft als Lenis bezeichnet und nicht als stimmhaft.	Gang [ˈgaŋ, g̊aŋ], Lager [ˈlaːgɐ, laːg̊ɐ]
h	stimmloser glottaler Frikativ	Haus [haʊs], Uhu [ˈuːhu]
j	Stimmhafter palataler Approximant	jung [jʊŋ], Boje [ˈboːjə]
k	stimmloser velarer Plosiv	Katze [ˈkat͡sə], Strecke [ʃtrɛkə]
l	stimmhafter lateraler alveolarer Approximant	Lamm [lam], alle [ˈalə]
m	stimmhafter bilabialer Nasal	Maus [maʊs], Dame [daːmə]
n	stimmhafter alveolarer Nasal	Nord [nɔrt], Kanne [ˈkanə]
ŋ	stimmhafter velarer Nasal	lang [laŋ], singen [ˈzɪŋən]
p	stimmloser bilabialer Plosiv	Pate [paːtə], Mappe [ˈmapə]
p͡f	stimmlose labiodentale Affrikate	Pfaffe [p͡fafə], Apfel [ˈap͡fəl]
r ʀ ʁ	stimmhafter alveolarer Vibrant ([r]), stimmhafter uvularer Vibrant ([ʀ]), stimmhafter uvularer Frikativ ([ʁ]) – Diese drei Laute sind freie Allophone. Ihre Verteilung ist lokal, wobei [r] fast ausschließlich in einigen südlichen Varietäten anzutreffen ist. Im Silbenauslaut wird das /r/ oft vokalisiert zu [ɐ̯], besonders nach langen Vokalen und in der unbetonten Endung /ər/, die bei Vokalisierung als [ɐ] relisiert wird.	rot [roːt, ʀoːt, ʁoːt], starre [ˈʃtarə, ˈʃtaʀə, ˈʃtaʁə], mit Vokalisierung: sehr [zeːɐ̯], besser [ˈbɛsɐ]
s	stimmloser alveolarer Frikativ	Straße [ˈʃtraːsə], Last [last], Fässer [ˈfɛsɐr]
ʃ	stimmloser postalveolarer Frikativ	Schule [ˈʃuːlə], Stier [ʃtiːr], Spur [ʃpuːr]
t	stimmloser alveolarer Plosiv	Tag [taːk], Vetter [ˈfɛtɐr]
t͡s	stimmlose alveolarer Affrikate	Zaun [t͡saʊn], Katze [ˈkat͡sə]
t͡ʃ	stimmlose postalveolare Affrikate	deutsch [dɔʏt͡ʃ], Kutsche [ˈkʊt͡ʃə]
v	stimmhafter labiodentaler Frikativ – Bisweilen wird dieser Laut als stimmhafter labiodentaler Approximant ([ʋ]) beschrieben.	Winter [ˈvɪntɐr], Löwe [løːvə]
x	stimmloser velarer Frikativ – Dieser Laut bildet zusammen mit [ç] ein komplementäres Allophon-Paar. Er tritt nach hinteren Vokalen auf (inklusive /a aː/). In nördlichen Varietäten erscheint er auch als Allophon /g/ im Silbenauslaut nach hinteren Vokalen (inklusive /a aː/.	lachen [ˈlaxən], nördliche Varietäten: sag [zaːx]
z	stimmhafter alveolarer Frikativ – Da dieser Laut in den südlichen Varietäten stimmlos ist ([z̥]), wird er oft als Lenis bezeichnet und nicht als stimmhaft.	sechs [zɛks, z̥ɛks], Wiese [ˈviːzə, ˈviːz̥ə]
ʒ	stimmhafter postalveolarer Frikativ – Dieser Laut tritt nur in Fremdwörtern auf. Da dieser Laut in den südlichen Varietäten stimmlos ist ([ʒ̊]), wird er oft als Lenis bezeichnet und nicht als stimmhaft.	Genie [ʒeˈniː, ʒ̊eˈniː], Plantage [planˈtaːʒə, planˈtaːʒ̊ə]

Table 4: German consonants and IPA expressions (Wikipedia, 2007a).

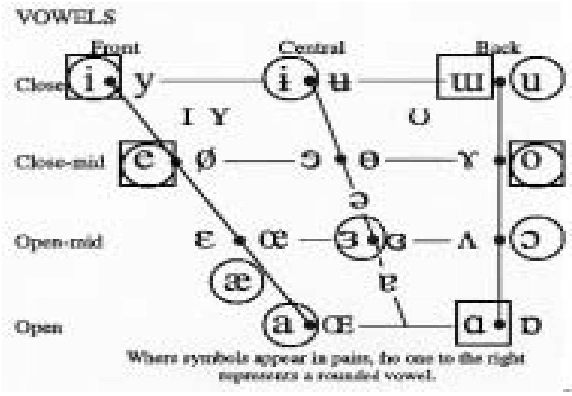

Figure 4: Japanese and Thai vowels on IPA (Retrieved from: Kasuriya, S. et al. 2002).[4]

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Table 5: Japanese and Thai consonants on IPA (Retrieved from: Kasuriya, S. et al. 2002).

3.3 Ontology of Phonemes

A natural way to present machine-processable concepts is by setting up an appropriate ontology. In the case of conceptual data for an automatic transcription system, the ontology must include the phonemes in a suitable representation and with a description; moreover, a grapheme which matches a given phoneme must be shown as well. In the project Electronic Metastructure for Endangered Languages Data, linguists are developing, among other things, an ontology of phonetics (EMELD, 2005; see Figure 5.) which can be used as a basis for the converter software.

A suitable method presenting the concepts which have to be processed by software is to provide an ontology comprising the classes, properties and instances of

[4] Thai phoneme positions on the IPA table are identified in the circles and the squares show the Japanese phoneme positions.

the field of interest. For an automatic transcription system the ontology must include the phonemes in a suitable representation and with a description; moreover, for all considered languages at least one grapheme which matches a given phoneme must also be included.

Figure 5: Section of the Phonetics ontology with front vowels and affricates.

This ontology shows the conceptualization of all IPA symbols (and some others also), but for the cross-language phoneme-grapheme matching process it lacks the presentation of suitable graphemes in different languages. An example showing how the ontology could be enhanced by typical graphemes is in Figure 6.

Template slots							
Slot name	**Documentation**	**Type**	**Allowed values/ classes**	**Position in syllable**	**Default**	**English**	**Thai**
rdfs:label	d-yogh ligature	*String*		begin		j	จ
rdfs:label	d-yogh ligature	*String*		end		j	จญ์

Figure 6: IPA Symbol "d-yogh ligature": Example of a class in the phonetics ontology extended by English and Thai graphemes.

3.4 Name Matching Algorithm – MetaSound

Methods which assume that the string representations capture sound are usually termed phonetic; however, it is important to note there may be no explicit phoneme structure present. A Yorkshireman trying to spell a name is not capturing the phoneme construction, due to his accent, but is spelling the name correctly but with his perception of the sounds represented by each letter, or syllable.

The language spelling of surnames is therefore at best an approximate phonetic representation. We define phonetic methods as an attempt to follow the sound structure present in the spelled ways since there can be no correct or standard spelling which is invariably accurate. In all the phonetic methods presented here, there is a mapping between particular spellings and a sound or sounds, and these mappings are usually captured as a set of transcription rules. Typical phonetic algorithms basically work by suppressing the vowel information (because it is unreliable) and giving the same code to letters or groups of letters which sound the same (e.g. PH sounds like F, so they are given the same code).

The main problem is of relating words which sound the same but are spelt differently; such words should have the same sound code. In the search for the most frequently used spelling of a particular name, the user would type that name, and the program would calculate its sound code, search the text(s) for all words with the same code, and present the user with the name which had the greatest frequency.

Snae et al (2007) have implemented a new phonetic base name matching technique called MetaSound (a combination of the Soundex and Metaphone algorithms), which is an improvement on Thai Soundex (Snae and Brueckner, 2007) and is used for finding name variants (spelling and phonetic variations) in the Thai Naming System, since Soundex and Metaphone (Snae, 2007) are phonetic coding algorithms which are designed primarily for use with English names only. Therefore, MetaSound is developed on the basis of commonplace rules of language pronounciation for matching words which sound and are spelt alike. The algorithms assign a value to a string based on the sounds. For example, people attempt to capture sound by writing down what they have heard and they believe that what they have written from what they heard is correct, e.g. “Smith” to “Smythe”. The following example is an implementation of MetaSound based on the rules of Thai pronunciation. The algorithm reduces the Thai alphabet to eight consonant sounds: K, D, B, NG, N, M, Y and W, as follows:

1. retain the first letter of the input name and this serves as the prefix letter
2. ignore vowels entirely
3. ignore mute pseudo-cluster combinations and letters with the mute indicator, called the Karan symbol (อ์)
4. If (While Loop) the input name is not converted to output MetaSound code and it is less than 5 characters, transcode the other letters to the MetaSound code as follows: letters and sounds are transcoded into digits.
   For example, for Thai letters:

letter	Sound	Letter is coded to:
ก ข ค ฅ ฆ	K	1
จ ฉ ช ซ ฌ ฎ ฏ ฑ ด ต ถ ท ธ ศ ส ษ	D	2
บ ป พ ฟ ผ ฝ ภ	B	3
ง	NG	4
น ล ฬ ร ณ ญ	N	5
ม	M	6
ย	Y	7
ว	W	8

5. output MetaSound code in the form letter, digit, digit, digit, digit
6. if output MetaSound code is less than five characters add trailing zeros otherwise drop the rightmost characters (the remaining letters are ignored)
7. return to MetaSound Code

For this, we can apply this MetaSound algorithm in relation to any language by changing and clustering groups of letters (of those languages which have the same phoneme) into one sound and one digit respectively. If the languages have more than nine sounds then we use "_" (underscore) to define 2 digits (or more) together, e.g. 1_0 (ten), 1_1 (eleven) and so on.

4. Architecture of the SPION System

The SPION system model/framework (Figure 7) works with syllable and pronunciation segmentation, an IPA representation, an ontology of phonemes and graphemes and a name-matching algorithm. The system expects a word or name from a source language either as manual input or as a retrieved string. With the help of a language-specific syllable segmentation and pronunciation model, the phonemes are extracted and converted into an IPA representation. At this point the word (or name) no longer has a meaning, but is rather a representation of the sounds which a native or near-native speaker would generate, using the grapheme. A different interpretation of this situation is that the item is no longer a lexeme but an allograph.

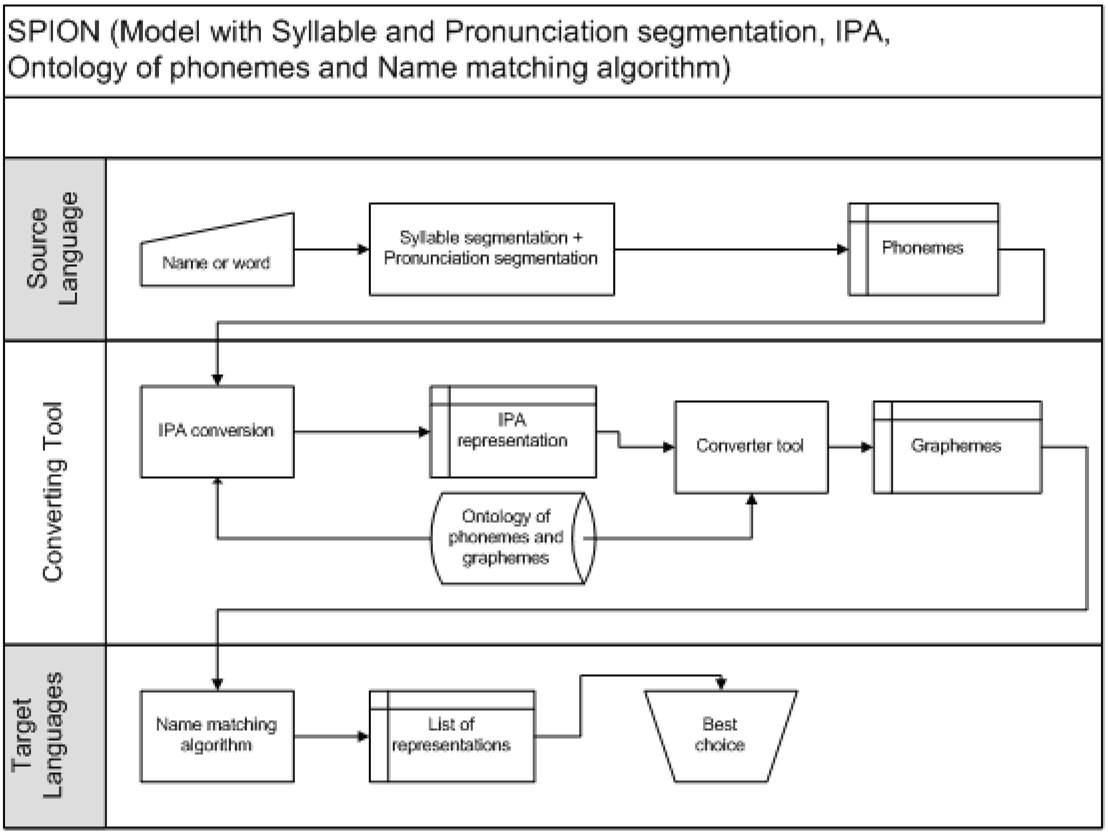

Figure 7: SPION system model/framework.

The next step is to convert the IPA representation of the word (or name) into a language-specific written version, represented by graphemes which native speakers of the target language can interpret in their own phonemic space. To coordinate this conversion process a phonetic ontology of phoneme-grapheme pairs is used, which shows for any given IPA symbol the best or nearest grapheme representation of the target languages which the system can cope with. The result of the converting tool is a phonetic characterization of a word (or name) of a source language in the writing system of the target language. When dealing with names, name-matching algorithms have proven to be useful in finding the closest matches between source languages and target languages. For names, SPION uses a novel name matching algorithm, called Meta-Sound. Meta-Sound generates a set of similar results for a grapheme representation in the target language and shows the nearest results in a list which a user can interpret and comment on (see Section 3.4 above).

5. Summary and Future Work

In this paper, we have described a framework for a multi-lingual transcription system named SPION. This framework integrates four advanced methods and tools: syllable and pronunciation segmentation, an IPA representation, an ontology of phonemes and graphemes and a name matching algorithm. In previous studies, we have implemented IT-TELLS and RESETT to transcribe English to Thai and Thai to English as part of the general project NARESUAN-M^2 (Name Recognition Expert System Using

Automated Name Matching Methods). The syllable and pronunciation model has been already been tested with the Thai dictionary and the accuracy is extremely high. The next step will be to implement the framework and test it with multi-lingual data.

References

Aroonmanakun, W. and W. Rivepiboon (2004) A Unified Model of Thai Word Segmentation and Romanization. In: Proceedings of The 18th Pacific Asia Conference on Language, Information and Computation, pp. 205–214, Tokyo, Japan.

Dematteis, K., Lutz, R., & McCallum-Bayliss, H. (1998) Whose name is it: Names, ownership and databases. Originally written for: 1998 Annual Meeting American Name Society San Francisco, CA.

EMELD (2005) A Phonetics Ontology, 2005. Available on-line from: http://emeld.net/workshop/2005/ontology-html/phonetics/ (accessed: 20 June 2007)

Gibbon, D. (2006) Can there be standards for spontaneous speech? Towards an ontology for speech resource exploitation, 2006. Available on-line from: http://wwwhomes.uni-bielefeld.de/~gibbon/Docs/gibbon-plenary-ontologiesforspeech02.pdf (accessed: 12 June 2007)

GOLD (2005) An Ontology for Descriptive Linguistics**,** 2005 Available on-line from: http://www.linguistics-ontology.org/gold.html (accessed: 20 June 2007)

Helpern, J. (2002) Lexicon-Based Orthographic Disambiguation in CJK Intelligent Inforation Retrieval. Proceedings of the 19th Conference on Computational Linguistics, COLING-2002, Taipei, Taiwan.

Helpern, J. (2006) Japanese Phomological Database (JPD), 2006 The CJK Dictionary Institute, Inc. Available on-line from: http://www.kanji.org/cjk/samples/jpd_e.htm (accessed: 20 June 2007)

Kamholz, D. (2005) An Ontology for Sounds and Sound Patterns, 2005. Available on-line from: http://emeld.org/workshop/2005/papers/kamholz-paper.doc (accessed: 6 May 2007)

Kasuriya, S., T. Jitsuhiro, G. Kikui, and Y. Sagisaka, Thai Speech Recognition by Acoustic Models Mapped from Japanese, Joint International Conference of SNLP-Oriental COCOSDA 2002. Available on-line from: http://vaja.nectec.or.th/view/paper/snlp2002_0091.pdf (accessed: 10 June 2007)

Knight, K. and J. Graehl (1998) 'Machine Transliteration'. *Computational Linguistics*. Vol. 24 (4).

Pinyin (2007) Available on-line from: http://www.pinyin.info/romanization/wadegiles/basic.html (accessed: 27 June 2007)

Qu, Y., G. Grefenstette and D. A. Evans (2003). Automatic transliteration for Japanese-to-English text retrieval, Annual ACM Conference on Research and Development in Information Retrieval. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 353–60.

Reaney, P. H. & Wilson, R. M. (1997). *A dictionary of English surnames.* Oxford: OUP.

Royal Thai General System of Transcription (1999) Available on-line from: http://www.royin.go.th/upload/246/FileUpload/416_2157.pdf (accessed: 2 May 2007)
Schultz, T., A. W. Black, S. Vogel, and M. Woszczcyna (2006) ‘Flexible Speech Translation Systems’. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol 14(2).
Sef, T. and M. Gams (2003) ‘Speaker (GOVOREC): A Complete Slovenian Text-to Speech System’. *International Journal of Speech Technology*, pp. 277–87.
Snae, C. (2007) ‘A Comparison and Analysis of Name Matching Algorithms’ *International Journal of Applied Science. Engineering and Technology*, Vol 4 no. 1, 252–57.
Snae, C. and M. Brueckner (2006) ‘Concept and Rule Based Naming System’. *The Information Universe: Journal of Issues in Informing Science and Information Technology*, Volume 3, 619–34.
Snae, C. and M. Bruecker (2007) A Semantic Web Integrated Searching System for Local Organizations with Thai Soundex, The 4th International Joint Conference on Computer Science and Engineering (JCSSE 2007) pp. 210–217, Khon Kean, Thailand.
Snae, C., K. Namahoot and M. Brueckner (2007) MetaSound: A New Phonetic Based Name Matching Algorithm for Thai Naming System, International Conference on Engineering, Applied Science and Technology (ICEAST 2007), Bangkok, Thailand.
Snae, C. and P. Pongcharoen (2007) Automatic Rule-based Expert System for English to Thai Transcription. The IASTED International Conference on Advance in Computer Science and Technology (ACST2007), pp. 342–47, Phuket, Thailand.
Snae, C. N. Singhadech, B. Emapana, and M. Brueckner (2006) Interactive Transliteration Tools for Explanation Level Language System (IT-TELLS), In Proceeding of ITC-CSCC 2006: The 21st International Technical Conference on Circuits/Systems, Computers and Communications, pp. 245–48, Chiang Mai. Thailand.
Wells, J. C. (1996) SAMPA for German, 1996. Available on-line from: http://www.phon.ucl.ac.uk/home/sampa/german.htm (accessed: 2 June 2007)
Wikipedia (2007) Available on-line from: http://en.wikipedia.org/wiki/IPA_chart_for_English (accessed: 10 June, 2007)
Wikipedia (2007a) Available on-line from: http://de.wikipedia.org/wiki/Aussprache_der_deutschen_Sprache (accessed: 15 June 2007)
Zheng, M., Q. Shi, W. Zhang and L. Cai (2005) Grapheme/to/Phoneme Conversion Based on a Fast TBL Algorithm in Mandarin TTS Systems. In: L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3614, 600–609.