

[*International Conference on the State of Machine Translation in America, Asia and Europe. Proceedings of IAI-MT86, 20-22 August 1986, Bürgerhaus, Dudweiler*]

Burhan Bukhari

BAPCOMP's Universal Computer System

Head Office:
34, East Rukneddine
Al Fayhan 3
Damascus, Syria

German Office:
Dr- J- Kizaoui
Am Tiergarten 1
6670 St. Ingbert

SUMMARY

=====

Name of the system	BAPCOMP'S UNIVERSAL COMPUTER SYSTEM
Status	Research
Type of system	MULTILINGUAL
Translated languages	GERMAN - FRENCH - ENGLISH - SPANISH- RUSSIAN- ARABISH
Speed of the system	Dependent on the used computer system
Type of analysis output	BAPCOMP'S SUPRALINGUA
Dictionaries	BAPCOMP'S NEW COMPILED DICTIONARY
Implementation language	Independent
Operating system	Any system usable
Inventor	Mr. BURHAN BUKHARI

BAPCOMP'S UNIVERSAL COMPUTER SYSTEM

1. TARGET

Based on long termed empirical researches, BAPCOMP started 1980 to develop the

"BAPCOMP'S UNIVERSAL COMPUTER SYSTEM"
according to "BUKHARI'S THEORY" for the purpose of

- PHONETICAL CHARACTER SET TRANSCRIPTION
- UNIVERSAL TYPESETTING
- MECHANICAL DIRECT TRANSLATION
- MECHANICAL SIMULTANEOUS TRANSLATION
(SPEECH SYNTHESIS)

applicable to most of languages in the world.

2. METHOD OF INVESTIGATION

In order to obtain this very complicated and revolutionary objectives resp. target first of all the languages concerned had to be analysed for the purpose to determine their very smallest basic particles being the last origin of written languages and their alphabets. It is obvious, that utilization of computers is more applicable the finer "FRAGMENTATION" of languages could be defined, since then numerous combinations of words are easily obtainable.

To reach this "FRAGMENTATION" resp. the required smallest particles of languages, the answers of many questions regarding the nature of languages concerned, their alphabets, their letters and their functions should be found and logically ordered.

Only for information, some of those questions to be investigated and answered could be mentioned hereinafter as follows:

- What is LANGUAGE ?
 - What is WRITING ?
 - How did WRITING start ?
 - How to construct resp. to build an ALPHABET for a hitherto UNWRITTEN LANGUAGE ?
 - Which effects does PRINTING have on WRITING ?
 - What is LETTER'S FUNCTION ?
 - Is it possible to deal SIMULTANEOUSLY with DIFFERENT LANGUAGES ?
 - Is it possible to work out a UNIVERSAL ALPHABET ?
 - What is the relationship between the LETTER appearing on the KEYBOARD and the LETTER stored within the PHOTOSetting MACHINE'S MEMORY ?
 - What is the relationship between the SHAPE of the LETTER and its PHONETIC VALUE ?
-
- Is it possible to work out a UNIVERSAL LANGUAGE ?

3. RESULTS

The FIRST_RESULT of several years of studies and investigations lead by Mr. BURHAN BUKHARI, the required

"BAPCOMP'S UNIVERSAL ALPHABET"

was achieved 1984, based on the phonetical reception and its aggregation of usual consonants, special consonants, complex consonants, vowels and syllables for more than forty languages. The validity of this "BAPCOMP'S UNIVERSAL ALPHABET" for further languages are still under development.

The "BAPCOMP'S UNIVERSAL ALPHABET" had been elaborated by Mr. B. BUKHARI under the assistance of a staff of linguists of several nationalities.

What we do mean by such an alphabet is, one that would contain all the diverse characteristics of all the forms of writing in all the languages of the world. And if we borrow the terminology

of mathematics, we can say:

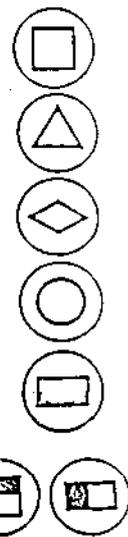
This alphabet could be the common denominator of all alphabets.

Functionally, this is a transkription alphabet whose role is to reduce the various forms of writing into standardized categories.

It is by no means a substitute for any language, its role being to a great extent as transcription as that of shorthand.

The "BAPCOMP'S UNIVERSAL ALPHABET" consists of special "BALNK AND SHADED GEOMETRICAL FIGURES" accompanied with "NUMERICAL SYMBOLS" according to the following scheme:

- Most frequent CONSONANTS in the languages concerned get the geometrical symbol of a SQUARE - BLANK
- SPECIAL CONSONANTS of particular languages get the geometrical symbol of a TRIANGLE - BLANK
- COMPLEX CONSONANTS of particular language get the geometrical symbol of RHOMBUS - BLANK
- VOWELS in the languages concerned get the geometrical symbol of a CIRCLE - BLANK
- DIACRETICA, FIGURES, MATHEMATICAL SYMBOLS...etc get the geometrical symbol of a RECTANGLE - BLANK
- SYLLABLES of particular languages get the corresponding above mentioned geometrical symbols, but in shaded shape



The alphabets of the developed languages had been put in relation with the "BAPCOMP'S UNIVERSAL ALPHABET" in a MATRIX, ordering vertically all letters of each alphabet and horizontally on the same line all similar letters of the alphabets of these languages. (See figure 1).

The languages which have already been studied, cover four groups of languages:

LATIN GROUP:	01 LATIN	02.FRENCH	03 ITALIAN	04 PORTUGUESE
	05 SPANISH	06 CATALAN	07 ROMANIAN	08 ENGLISH
	09 GERMAN	10 DUTCH	11 SWEDISH	12 NORWEGIAN
	13 DANISH	14 ICELANDIC	15 CZECH	16 POLISH
	17 FINNISH	18 HUNGARIAN	19 ESPERANTO	20 SEBRO-CROAT
	21 TURKISH	22 ALBANIAN	23 WELSH	24 SWAHILI
	25 MALTESE	26 INDONESIAN	27 SAMOAN	

CYRILLIC GROUP:	28 RUSSIAN	29 UZBEK	30 BULGARIAN
	31 UKRAINIAN	32 USTEIN	
ARABIC GROUP :	33 ARABIC	34 PERSIAN	35 URDU
	36 PUSHTU	37 KURDISH	38 UZBEK
SEPERATES GR. :	39 HEBREW	40 ARMENIAN	41 GEORGIAN
	42 SYRIAC	43 GREEK	44 AMHARIC

The INDIAN LANGUAGES are still under development.

The SECOND RESULT of these researches and investigations was the construction of the

"BAPCOMP'S UNIVERSAL KEYBOARD"

The keys of this keyboard do not carry any more letters but "GEOMETRICAL and NUMERICAL SYMBOLS"; the keyboard has the standard dimensions of usual keyboards. (See figure 2). Accordingly the "BAPCOMP'S UNIVERSAL KEYBOARD" operates with the "BAPCOMP'S UNIVERSAL ALPHABET".

With this keyboard a number of about 1700 symbols per shift can be achieved.

The "BAPCOMP'S UNIVERSAL KEYBOARD" enables the transfer of the "BAPCOMP'S UNVERSAL ALPHABET" into a standard computer system by using the

"BAPCOMP'S DOUBLE TOUCH SYSTEM"

The "BAPCOMP'S DOUBLE TOUCH SYSTEM" means that two keys must be pressed simultaneously for the in put of one letter, diacretic,... ..etc, namely one key for geometrical symbol and one for numerical symbol. (See figure 1).

The THIRD RESULT of these researches and investigations was the approval of the practical application of

"BAPCOMP'S UNIVERSAL ALPHABET"

"BAPCOMP'S UNIVERSAL KEYBOARD"

"BAPCOMP'S DOUBLE TOUCH SYSTEM"

based on the elaborated tables for the concerned languages.

This practical approval has easily shown the practicability
BUKHARI'S THEORY by

- REPRODUCTION OF THE IN PUT SYMBOLS INTO
THE CORRESPONDING ALPHABET (see figure 3).
- PHONETICAL CHARACTER SET TRANSCRIPTION
(see figure 4)
- UNIVERSAL TYPESETTING

This phase of BAPCOMP's researches and investigations allows to
cover practically all following application fields, i.e.

- LIBRARIES
- MUSIC HOUSES
- UNIVERSITIES
- SCIENTIFIC INSTITUTIONS
- PRINTING COMPANIES
- PUBLISHING COMPANIES
- ADVERTISING COMPANIES
- TRADING COMPANIES
- COMMUNICATIONS
- LANGUAGE TRADING CENTRES
- SPEECH SYNTHESIS
-

The FOURTH RESULT was achieved, after the successful approval of the afore mentioned phase, by finding the

"BAPCOMP'S SUPRALINGUA"

which is the basis of the

"BAPCOMP'S UNIVERSAL MECHANICAL DIRECT TRANSLATION SYSTEM"

The "BAPCOMP'S SUPRALINGUA" had been obtained according to "BUKHARI'S THEORY" by fragmentation the language systems into their smallest components which vary from language to language. The "BAPCOMP'S SUPRALINGUA" considers all additional and variations used within the concerned languages too.

The afore mentioned fragmentation also applies to language pattern to which the seemingly endless variety of structures exhibited by language can be reduced.

The "BAPCOMP'S SUPRALINGUA" allows to translate any language to any other.

The FIFTH RESULT of BAPCOMP'S researches and investigations was achieved by the practical approval no. 1 by using the "BAPCOMP'S SUPRALINGUA" for testing the afore mentioned

"BAPCOMP'S UNIVERSAL MECHANICAL DIRECT TRANSLATION SYSTEM"

This first implementation test was made for two LATIN languages namely, ENGLISH - FRENCH and FRENCH - ENGLISH using a PERSONAL COMPUTER with compact floppy disc.

The data for this test consist of three verbs, namely, WRITE - BRING - READ, resp. ECRIVER - APPORTER - LIRE taking into consideration all grammatical combinations with all personal pronouns.

The current software is a FIRST ATTEMPT for automatic translation.

Therefore, the program requires some extra pieces of information to resolve morpho-syntactic ambiguity of the verbs.

The tenses are identified by specific numbers pointing out the tense implication for the target languages.

Example: read - infinitive
Read1 - present form
Read2 - past form
Read3 - past participle

Example for the FRENCH equivalent of the
ENGLISH tenses:

Il lit 1 - present simple
Il lit 2 - present continuous
Il lit 3 - present perfect continuous

Consequently 168 combinations can be obtained by this very simple program for each language. Some of those combinations are shown in figure 5.

Since the afore mentioned combinations of sentences were not sufficient enough in order to judge the practicability of the "BAPCOMP'S UNIVERSAL MECHANICAL DIRECT TRANSLATION SYSTEM" another software program no. 2 had been prepared which enables us to obtain about 50,000 combinations of sentences for each language, ENGLISH - FRENCH resp. FRENCH - ENGLISH.

The program no. 2 needs some special information as follows:

1. VERBS : READ - WRITE - BRING, resp. LIRE - ECRIVER - APPORTER
2. NOUNS : BOOK - LETTER - ARTICLE - REPORT - STORY resp.
LIVRE- LETTRE - ARTICLE - RAPPORT- HISTOIRE
in singular and pluriel form
3. ARTICLE : A - AN - THE resp.
UN- UNE- DES - LE - LA - LES
4. POSSESSIVE
ADJECTIVES: MY - OUR - YOUR - HIS - HER - ITS - THEIR resp.
MA - MON - MES - TON - TA - TES
SA - SON - SES - NOTRE - NOS - VOTRE
VOS - LEUR - LEURS

5. SPECIAL INFORMATION:

READ	- infinitive, present
READ2	- past and past participle
YOU	- 2nd person singular
YOU2	- 2nd person plural
THEY	- masculine
THEY2	- feminine

for the FRENCH equivalents:

IL LIT	- present simple
IL LIT 2	- present continuous
IL LIT 3	- present perfect continuous
IL A LU	- simple past
IL A LU 2	- past perfect

Some of those combinations are shown in figure 6.

BAPCOMP is just preparing the implementation no. 3 taking six languages into account, namely

"ENGLISH - FRENCH - GERMAN - SPANISH - RUSSIAN - ARABIC"

This software installed on an IBM PC shall enable us to obtain about 20 mio combinations resp. sentences. The user shall not give any extra information with those words, which are ambiguous, since the program itself has to find the word required.

The result of the software program no. 3 shall be demonstrated at the conference organized by IAI/EUROTRA-D.

4. SUMMARY

Based on BUKHARI'S THEORY, BAPCOMP could work out within the last six years under the assistance of a staff of linguists of several nationalities all basic elements and programmes needed for the

"BAPCOMP'S UNIVERSAL COMPUTER SYSTEM"

for phonetical character set transkription and mechanical direct translation which are:

" BAPCOMP'S UNIVERSAL ALPHABET"
 " BAPCOMP'S UNIVERSAL KEYBOARD"
 " BAPCOMP'S SUPRALINGUA"

FIGURE 1 : SAMPLES OF BAPCOMP'S ALPHABET-MATRIX
 BASED ON BUKHARI'S THEORY
 FOR SEVERAL LATIN LANGUAGES, ARABIC
 HEBREW AND AMHARIC

LATIN اللاتينية

Computer Code	Phonetic Symbols	اللاتينية LATIN	الفرنسية FRENCH	الإيطالية ITALIAN	الإسبانية SPANISH	البرتغالية PORTUGUESE	الكاتالانية CATALAN	الرومانية ROMANIAN	الانجليزية ENGLISH	الالمانية GERMAN	الهولندية DUTCH	السويدية SWEDISH
○ 1	Aa	Á á	À à	À à	À à	À à	À à	À à	À à	À à	À à	À à
○ 2	Ii	Í í	Î î	Î î	Î î	Î î	Î î	Î î	Î î	Î î	Î î	Î î
○ 3	Ee	É é	È è	È è	È è	È è	È è	È è	È è	È è	È è	È è
○ 4	Uu	Ú ú	Ù ù	Ù ù	Ù ù	Ù ù	Ù ù	Ù ù	Ù ù	Ù ù	Ù ù	Ù ù
○ 5	Oo	Ó ó	Ò ò	Ò ò	Ò ò	Ò ò	Ò ò	Ò ò	Ò ò	Ò ò	Ò ò	Ò ò
○ 6	À à	Á á			À à	À à	À à					

ARABIC العربية

Computer Code	Phonetic Symbols											
1	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب	ب
2	پ	پ	پ	پ	پ	پ	پ	پ	پ	پ	پ	پ
3	د	د	د	د	د	د	د	د	د	د	د	د
4	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت
5	ث	ث	ث	ث	ث	ث	ث	ث	ث	ث	ث	ث
6	ي	ي	ي	ي	ي	ي	ي	ي	ي	ي	ي	ي

HEBREW العبرية

Computer Code	Phonetic Symbols											
○ 1	ב	ב	ב	ב	ב	ב	ב	ב	ב	ב	ב	ב
○ 2	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ
○ 3	ד	ד	ד	ד	ד	ד	ד	ד	ד	ד	ד	ד
○ 4	ת	ת	ת	ת	ת	ת	ת	ת	ת	ת	ת	ת
○ 5	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ
○ 6	י	י	י	י	י	י	י	י	י	י	י	י

AMHARIC الامهرية

Computer Code	Phonetic Symbols											
○ 1	ב	ב	ב	ב	ב	ב	ב	ב	ב	ב	ב	ב
○ 2	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ
○ 3	ד	ד	ד	ד	ד	ד	ד	ד	ד	ד	ד	ד
○ 4	ת	ת	ת	ת	ת	ת	ת	ת	ת	ת	ת	ת
○ 5	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ	פ
○ 6	י	י	י	י	י	י	י	י	י	י	י	י

FIGURE 2 : BAPCOMP'S UNIVERSAL KEYBOARD"

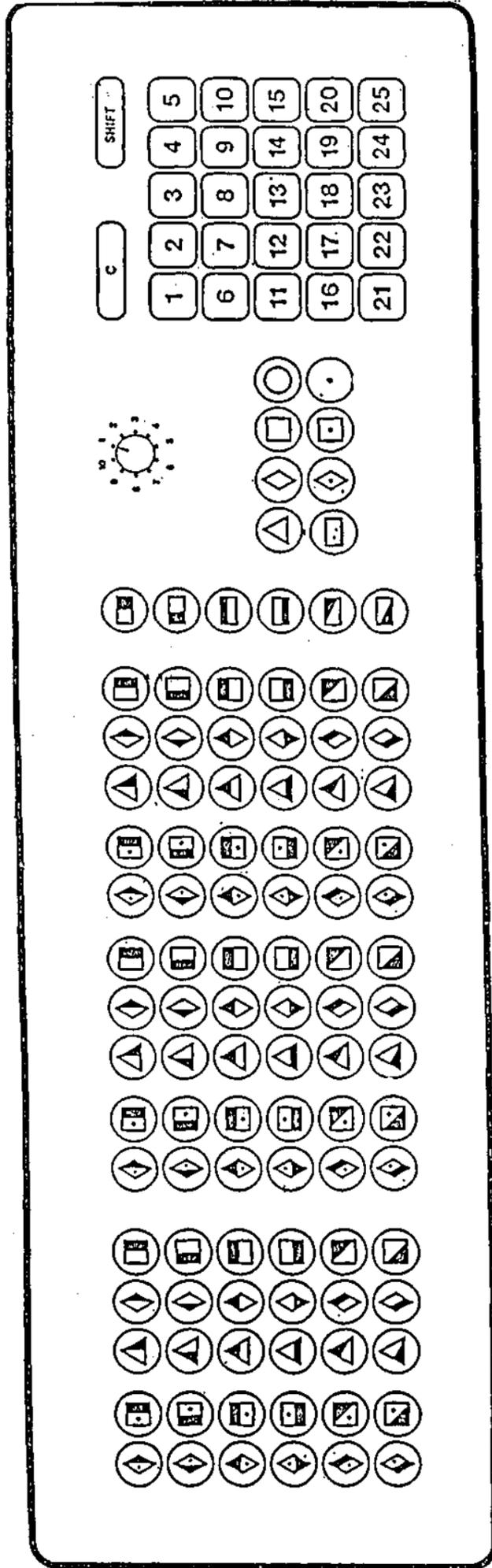


FIGURE 3 : REPRODUCTION OF THE IN PUT SYMBOLS
 INTO THE CORRESPONDING ALPHABETS
 BY USING BAPCOMP'S UNIVERSAL KEYBOARD

LATIN LANGUAGES

B P D Ď T ě V G Ğ J Ĵ H Ĥ X K C Q M N Ń L Ł R Ŕ S Ś Š †
 Š Š Z Ž Ž Ź W Ū Y Ÿ Ŷ B Ç Ć Ğ ħ †
 A Å Ä Å Æ E E É È Ě Ē Ę I İ Í Î Ï ÿ †
 O Ó Ô Õ Ö Ø Œ U Ů Ū Ŭ Ů Ū Ŭ †
 A

TAMILI - LANGUAGE

சிறுவன் அவனது தாயார் †
 சாவித்திரி தேவியுடன் †
 பஸ்சில் பயணம் செய் †
 தான் . தாயும் , மகனும் †
 பஸ்சில் இருந்து இறங்கும் †
 போது , வேறு இரண்டு †
 பெண்களும் கூடவே †
 இடித்துக்கொண்டு †
 இறங்கினார்கள் . கண் †
 இமைக்கும் நேரத்தில் †
 சாவித்திரி தேவியின் †
 கழுத்திலிருந்து தங்கச் †
 சங்கிலி யை பறித்துக் †
 கொண்டு அந்த இரண்டு †
 பெண்களும் தலைதெறிக்க †
 ஓடினார்கள் . †
 டீதைப் பார்த்ததும் †
 சிறுவன் யஸ்தீஷ் †
 அந்த பெண்களை பின் †
 தொடர்ந்து விரட்டிச் †
 சென்றான் . சிறிது தூரம் †
 விரட்டிச் சென்ற பிறகு , †
 அவர்களில் ஒரு பெண்ணை †
 அவன் பிடித்து விட் †
 டான் . அப்போது அந்தப் †
 பருதியில் காவலுக்கு †
 நின்றது கொண்டிருந்த †

FIGURE 4 : PHONETICAL CHARACTER SET TRANSCRIPTION
FROM LATIN SPELL LANGUAGES TO RUSSIAN
BY USING BAPCOMP'S UNIVERSAL KEYBOARD

Bukhari phonetic keyboard†

Set of words in different languages demonstrating the †
flexibility of the stored character set. †

†

leō --- leāo --- löwe --- löve --- oroszlān †

Monday --- Montag --- dimanche --- domenica --- domingo †

†

Букҳари пҳонэти кэйбоард†

Лэг оф ордс ин диффэрэнт лангуагэс дэмонстратинг тхэ †
флэуибилитй оф тхэ сторэд ҳара тэр сэт †

†

лэ лэ о лй э лувэ
Кондай Контаг дитан хэ

оросэл н †

домэни а

доминго †

FIGURE 6A: BAPCOMP'S UNIVERSAL MECHANICAL DIRECT
TRANSLATION SYSTEM BY USING THE
BAPCOMP'S SUPRALINGUA BASED ON
BUKHARI'S THEORY
Program no. 2

ENGLISH - FRENCH

i read, her books	je lis ses livres
you read their letters	tu lis leurs lettres
he reads his story	il lit son histoire
we read articles	nous lisons des articles
they read reports	ils lisent des rapports
we bring the books	nous apportons les livres
we are bringing the letters	nous apportons les lettres
they had been bringing a report	ils apportaient un rapport
we shall write stories	nous écrirons des histoires
they will be bringing the books	ils apporteront les livres
she does not write a book	elle n'écrit pas un livre
iam not reading his report	je ne lis pas son rapport
we read articles	nous lisons des articles
idid not write her book	je n'ai pas écrit son livre
i had not been reading a book	je ne lisais pas un livre
i do not read letters	je ne lis pas des lettres
they had brought a book	ils avaient apporte un livre
you2 read their letters	vous lisez leurs lettres
they2 read reports	elles lisent des rapports

FIGURE 6B: BAPCOMP'S UNIVERSAL MECHANICAL DIRECT
 TRANSLATION SYSTEM BY USING THE
 BAPCOMP'S SUPRALINGUA BASED ON
 BUKHARI'S THEORY
 Program no. 2

FRENCH - ENGLISH

je lis une lettre	i read a letter
elle lit son histoire	she reads his story
nous lisons des articles	we read articles
elles lisent des livres	they read books
j' écris la lettre	i write the letter
j' avais apporte des rapports	i had brought reports
elle a lu un livre	she read a book
il lira une lettre	he will read a letter
elle aura lu des articles	she will have read articles
nous ne lisons pas une lettre	we do not read a letter
elle n' avait pas lu un livre	she had not read a book
il ne lira pas une lettre	he will not read a letter
elle n' aura pas lu un livre	she will not have read a book
elle lit un livre 2	she is reading a book
elle lit un livre 3	she has been reading a book
elle lisait une lettre	she was reading a letter
elle lisait une lettre 2	she had been reading a letter
elle lira un livre 2	she will be reading a book
elle lira un livre 3	she will have been reading a bo
elle ne lira pas un livre 2	she will not be reading a book
je ne lirai pas un livre	i shall not read a book
je ne lirai pas un livre 3	i shall not have been reading a book