

Catherine Ray

Dr. Muffarah Jahangeer, Dr. Joseph Marr

## Introduction to CAMEL

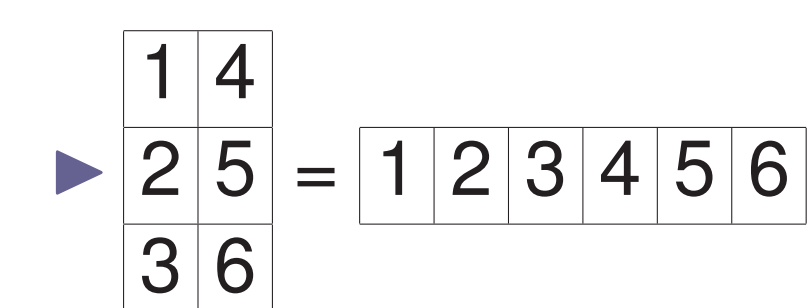
- ▶ CAMEL is an acronym of **C**ontextu**A**I **M**achin**E** Learning, it is a machine learning program that uses Braille as a language platform.
- ▶ uses context of unknown symbols to deduce meaning and compress information.
- ▶ provided the meaning of an initial set of symbols (a dictionary, or **dict**). CAMEL deduces meanings of unknowns and adds these meanings to the dict.
- ▶ grows more accurate as the dict increases in size and options. Some symbols differ in meaning depending on their context. These translation options are stored in the dict in the form of Map[String, TranslationOptions].

## What is Grade 2 Braille?

- ▶ As English words are composed of letters, Braille words are composed of Braille cells.
- ▶ **Contractions** are special characters used to reduce the length of words.
- ▶ Some contractions stand for a whole word.  
For example: 'for' = ⠠; 'and' = ⠞; 'the' = ⠠.
- ▶ Other contractions stand for a group of letters within a word. In the example below, the contraction 'ing' is used in the word 'sing' and as an ending in the word 'playing.'  
{ing} = ⠠; 's' + {ing} = ⠠⠠; 'play' + {ing} = ⠠⠠⠠.
- ▶ Grade 1 Braille is uncontracted Braille.
- ▶ Grade 2 Braille consists of Grade 1 Braille symbols and additional contracted cells.

## Binary Braille

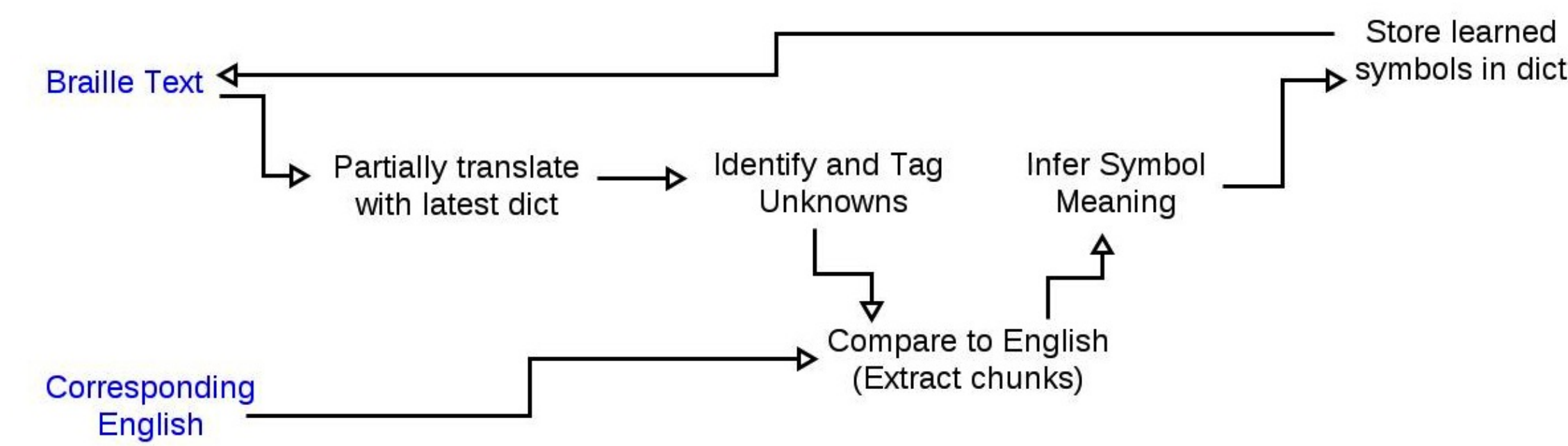
- ▶ The Braille alphabet is depicted by a cell that contains six raised/flat dots, numbered one through six beginning with the dot in the upper left-hand corner with the number descending the columns (see figure below).
- ▶ Let "0" = flat, "1" = raised. The 3x2 matrix (Braille cell) is represented as a 1x6 bitstring (Binary Braille).



▶ Thus, "c" = ⠠ ≡  $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \equiv \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \equiv 100100 \equiv 100100$

## String Processing Method

CAMEL deduces the complex grammar rules of Grade 2 Braille given partially translated text.



CAMEL learns new symbols by taking 2 input text files (Braille text and corresponding English text), and analyzing them until all unknowns are identified, their meanings are found, and said symbols and their meanings are added to the dictionary.

## Methods of Tagging and Text Extraction

CAMEL must *Tag Unknowns* & *Compare to English(Extract Chunks)* to infer symbol meaning. Four different tag types were used: end, front, mid, and full-word. Below are examples of how these different types of tags were each used to extract meaning.

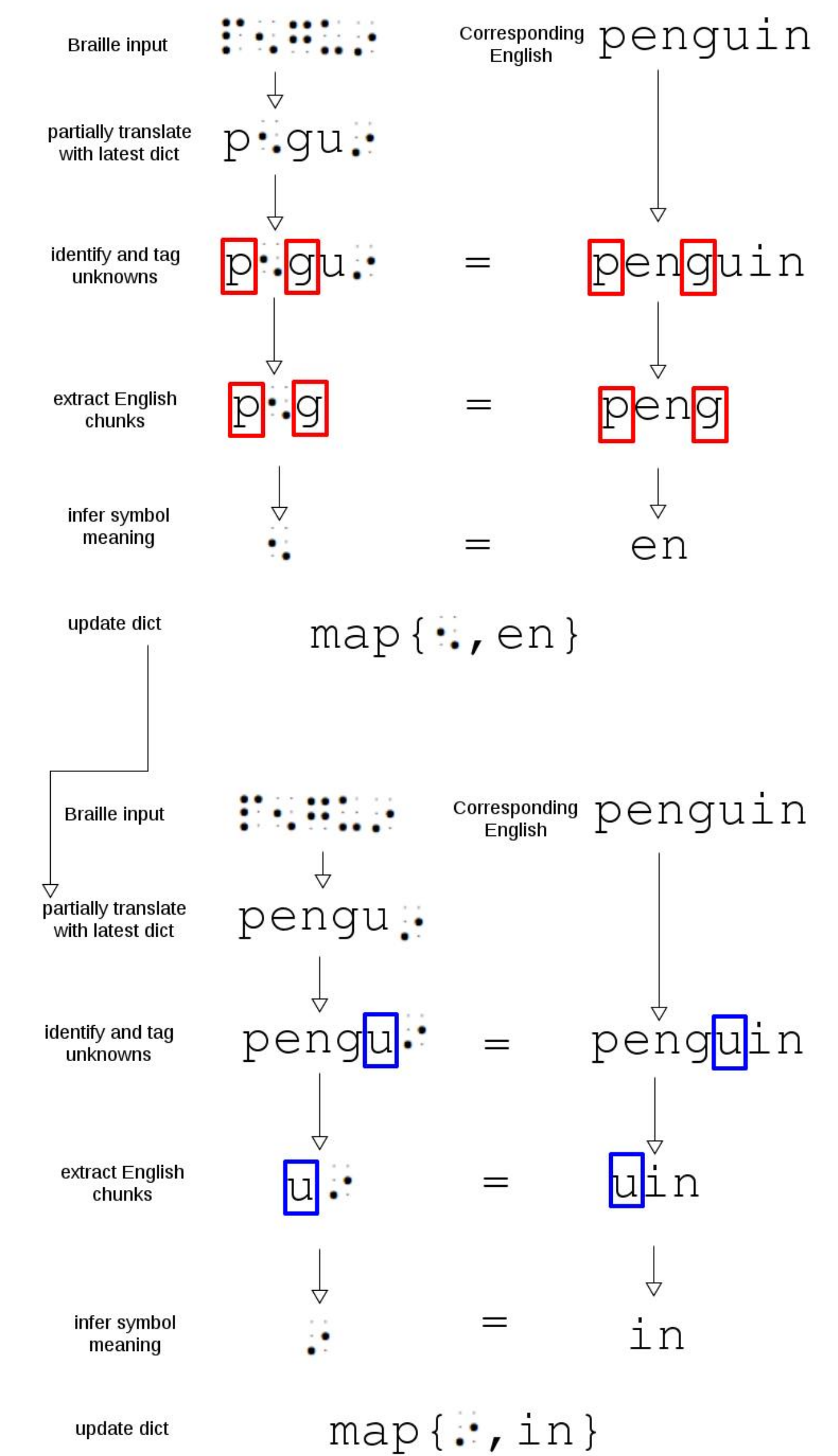
Tag Types	partially translated Braille example	regular expression used to extract tag	resulting tag	corresponding English match	chunk remaining after tag removal
end	off ⠠	$\backslash w^* (?=\backslash d)$	off	offer	er
front	⠠side	$[a-z]^+$	side	inside	in
mid (2 steps)	qu⠠ch	$\backslash w^* (?=\backslash d)$	qu	quench	<del>ench</del>
	⠠ch	$[a-z]^+$	ch	ench	en
full-word (alternate translation)	p	None	None	people	people
(unknown symbol)	⠠	None	None	for	for

## Acknowledgements

I'd like to thank: Dr. Marr for introducing me to the beautiful complexity of Computational Semantics; Dr. Jahangeer for her amazing encouragement and her willingness to answer questions; Dr. Borne, Dr. Marr, and Dr. Papaconstantopoulos for supporting me and convincing me to major in Computational and Data Sciences; www.braillebookstore.com for printing and distributing books in Grade 2 Braille, excerpts of which helped train CAMEL.

## Using Contracted Braille As a Platform

An example of this process infers the symbols that represent *en* and *in* using the word *penguin* (contracted to *p{en}gu{in}* in Grade 2 Braille).



## Results and Conclusions

### Safety of Community

- ▶ commercial application in development that will prevent future mislabeling, such as



- ▶ allows sighted people to protect the blind community

### Proof of Concept

- ▶ 1<sup>st</sup> successful automated program that learns compressed Braille
- ▶ translation system is effective for arbitrary symbol systems
- ▶ language platform easily changed