From Berger, Della Pietra, and Della Pietra (1996).

Table 2
Most frequent French translations of *in* as
estimated using EM-training. (OTHER)
represents a catch-all classifier for any
French phrase not listed, none of which
had a probability exceeding 0.0043.

Translation	Probability
dans	0.3004
à	0.2275
de	0.1428
en	0.1361
pour	0.0349
(OTHER)	0.0290
au cours de	0.0233
	0.0154
sur	0.0123
par	0.0101
pendant	0.0044

Next we define the set of candidate features. For this application, we employ features that are indicator functions of simply described sets. Specifically, we consider functions $f(x,y)$ that are one if y is some particular French word and the context x contains a given English word, and are zero otherwise. We employ the following notation to represent these features:

$$f_1(x,y) = \begin{cases} 1 & \text{if } y = en \text{ and } April \in \boxed{\ \ |\ \ |\ \ |\ \bullet\ |\ \ |\ \ } \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(x,y) = \begin{cases} 1 & \text{if } y = pendant \text{ and } weeks \in \boxed{\ \ |\ \ |\ \ |\ \bullet\ |\ \bullet\ |\ \bullet\ } \\ 0 & \text{otherwise} \end{cases}$$

Here $f_1 = 1$ when *April* follows *in* and *en* is the translation of *in*; $f_2 = 1$ when *weeks* is one of the three words following *in* and *pendant* is the translation.

Table 5
Maximum entropy model to predict French translation of *in*. Features shown here were the first features selected not from template 1. [*verb marker*] denotes a morphological marker inserted to indicate the presence of a verb as the next word.

	Feature $f(x,y)$	$\sim\Delta L(\mathcal{S},f)$	$L(p)$
y=à and *Canada* $\in$	[| | | • | |]	0.0415	-2.9674
y=à and *House* $\in$	[| | | • | |]	0.0361	-2.9281
y=en and *the* $\in$	[| | | • | |]	0.0221	-2.8944
y=pour and *order* $\in$	[| | | • | |]	0.0224	-2.8703
y=dans and *speech* $\in$	[| | | • | • | •]	0.0190	-2.8525
y=dans and *area* $\in$	[| | | • | • | •]	0.0153	-2.8377
y=de and *increase* $\in$	[• | • | • | | |]	0.0151	-2.8209
y=[verb marker] and *my* $\in$	[| | | • | |]	0.0141	-2.8034
y=dans and *case* $\in$	[| | | • | • | •]	0.0116	-2.7918
y=au cours de and *year* $\in$	[| | | • | • | •]	0.0104	-2.7792