

Unesco translationum

October 8, 2014

This notebook provides a detailed view of data from (UNESCO)[<http://www.unesco.org/xtrans/bsstatexp.aspx>] regarding the existing translations from books from one language into another. These results are based on data downloaded on 6 october 2014. Downloaded files have been slightly modified in order to be readable using python-pandas package.

```
In [1]: import pandas as pd
import ipy_table as pytbl
```

```
In [2]: %pylab inline
```

Populating the interactive namespace from numpy and matplotlib

```
In [3]: data = pd.read_csv('unesco_xtrans_stats.csv', sep='\t', index_col=0)
data = data.fillna(0)
```

1 Translations from English

To begin with, we examine the total number of books originally written in English and which have been translated into another language. The table below indicates also the total number of languages into which English books have been translated.

```
In [4]: data_en = data['English']
total_en = '{:,.0f}'.format(data_en.sum())
langs_en = '{:,.0f}'.format(len(data_en[data_en>0]))

table = []
table.append(['English', '' ])
table.append(['Total files', total_en ])
table.append(['Translated to languages', langs_en])

pytbl.make_table(table)
pytbl.set_cell_style(0,0, bold=True, column_span=2, color='gray')
pytbl.set_column_style(1, align='right')
```

Out[4]:

English	
Total files	1,267,868
Translated to languages	367

The following chart illustrates the percentage of books translated from English into other languages. Notice that languages which account for less than 2.5% of the total translations have been aggregated together as 'others'.

```
In [5]: pct = 0.025
data_en_pie = data_en[data_en>0]
```

```

data_en_pie = data_en_pie / data_en_pie.sum()
data_en_pie.sort()
fracs = data_en_pie[data_en_pie>pct].values.tolist()
labels = data_en_pie[data_en_pie>pct].index.tolist()
fracs.append(1 - sum(frac))
labels.append('Other')

cs = cm.Set1(np.arange(len(labels))/len(labels))
pie(frac, labels=labels, autopct='%1.1f%%', shadow=True, startangle=270, colors=cs)
title('Languages English books\nare translated to\n')
axis('equal');

```



2 Translations into English

Conversely, we examine the number of books which get translated from other languages into English. The following table indicates the total number of translations into English and from how many languages these translations occur.

```

In [115]: data_2_en = data.transpose()['English']
total_2_en = '{:,.0f}'.format(data_2_en.sum())
langs_2_en = '{:,.0f}'.format(len(data_2_en[data_2_en>0]))

table = [];
table.append(['English', '' ])
table.append(['Total files', total_2_en ])

```

```

table.append(['Translated from languages' , langs_2_en])

pytbl.make_table(table)
pytbl.set_cell_style(0,0, bold=True, column_span=2, color='gray')
pytbl.set_column_style(1, align='right')

```

Out[115]:

English	
Total files	167,289
Translated from languages	495

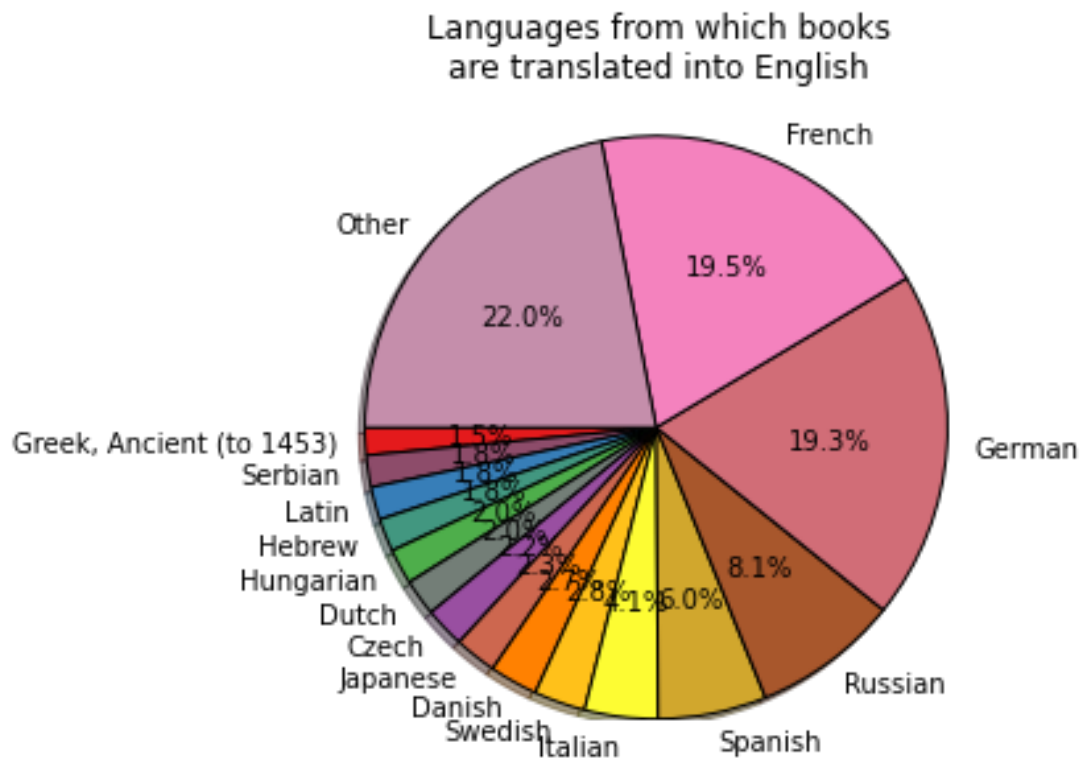
The following chart indicates the percentage of translations from other languages into English. Notice that Dutch features in the top 10 languages translated into English.

```

In [122]: pct = 0.015
data_en_pie = data_2_en[data_2_en>0]
data_en_pie = data_en_pie / data_en_pie.sum()
data_en_pie.sort()
fracs = data_en_pie[data_en_pie>pct].values.tolist()
labels = data_en_pie[data_en_pie>pct].index.tolist()
fracs.append(1 - sum(frac))
labels.append('Other')

cs = cm.Set1(np.arange(len(labels))/len(labels))
pie(frac, labels=labels, autopct='%1.1f%%', shadow=True, startangle=180, colors=cs)
title('Languages from which books\nare translated into English\n')
axis('equal');

```



3 Translations from Dutch

Next, we examine the total number of books originally written in Dutch and which have been translated into another language. The table below indicates also the total number of languages into which Dutch books have been translated.

```
In [6]: data_nl = data['Dutch']
        total_nl = '{:,.0f}'.format(data_nl.sum())
        langs_nl = '{:,.0f}'.format(len(data_nl[data_nl>0]))

        table = []
        table.append(['Dutch', ''])
        table.append(['Total files', total_nl])
        table.append(['Translated to languages', langs_nl])

        pytbl.make_table(table)
        pytbl.set_cell_style(0,0, bold=True, column_span=2, color='gray')
        pytbl.set_column_style(1, align='right')
```

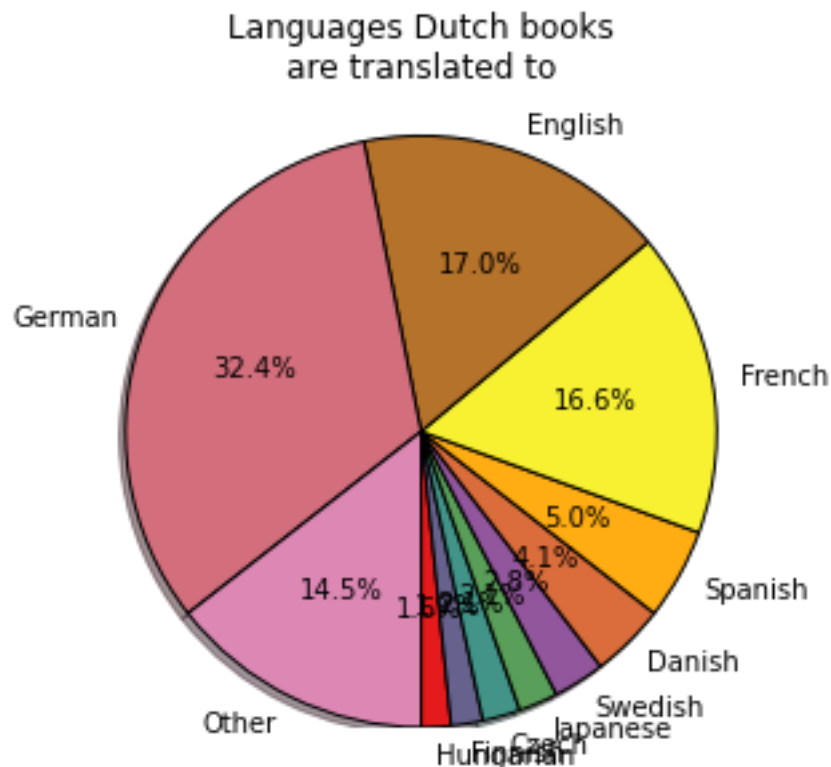
Out[6]:

Dutch	
Total files	19,907
Translated to languages	78

The following chart illustrates the percentage of books translated from Dutch into other languages. Notice that languages which account for less than 1.5% of the total translations have been aggregated together as 'others'.

```
In [7]: pct = 0.015
        data_nl_pie = data_nl[data_nl>0]
        data_nl_pie = data_nl_pie / data_nl_pie.sum()
        data_nl_pie.sort()
        fracs = data_nl_pie[data_nl_pie>pct].values.tolist()
        labels = data_nl_pie[data_nl_pie>pct].index.tolist()
        fracs.append(1 - sum(fracs))
        labels.append('Other')

        cs = cm.Set1(np.arange(len(labels))/len(labels))
        pie(fracs, labels=labels, autopct='%1.1f%%', shadow=True, startangle=270, colors=cs)
        title('Languages Dutch books\nare translated to\n')
        axis('equal');
```



4 Translations into Dutch

Conversely, we examine the number of books which get translated from other languages into Dutch. The following table indicates the total number of translations into Dutch and from how many languages these translations occur.

```
In [123]: data_2_n1 = data.transpose()['Dutch']
total_2_n1 = '{:,.0f}'.format(data_2_n1.sum())
langs_2_n1 = '{:,.0f}'.format(len(data_2_n1[data_2_n1>0]))

table = [];
table.append(['Dutch', ''])
table.append(['Total files', total_2_n1 ])
table.append(['Translated from languages', langs_2_n1 ])

pytbl.make_table(table)
pytbl.set_cell_style(0,0, bold=True, column_span=2, color='gray')
pytbl.set_column_style(1, align='right')
```

Out[123]:

Dutch	
Total files	111,713
Translated from languages	133

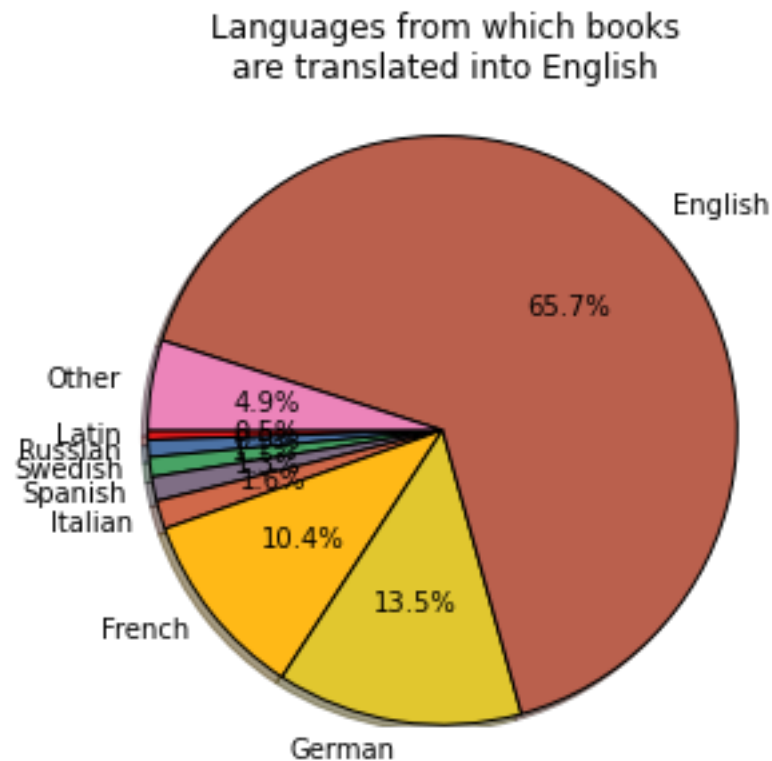
The following chart illustrates the percentage of translations from other languages into Dutch. Notice that English represents by far the majority of books which get translated into Dutch.

```

In [139]: pct = 0.005
data_nl_pie = data_2_nl[data_2_nl>0]
data_nl_pie = data_nl_pie / data_nl_pie.sum()
data_nl_pie.sort()
fracs = data_nl_pie[data_nl_pie>pct].values.tolist()
labels = data_nl_pie[data_nl_pie>pct].index.tolist()
fracs.append(1 - sum(frac))
labels.append('Other')

cs = cm.Set1(np.arange(len(labels))/len(labels))
pie(frac, labels=labels, autopct='%1.1f%%', shadow=True, startangle=180, colors=cs)
title('Languages from which books\nare translated into English\n')
axis('equal');

```



5 Between English and Dutch

And the following table indicates the number of books which have been published from English to Dutch and conversely from Dutch to English.

```

In [8]: nl_en = '{:,.0f}'.format(data['Dutch']['English'])
en_nl = '{:,.0f}'.format(data['English']['Dutch'])

table = [];
table.append(['', '', 'To', ''])
table.append(['', '', 'English', 'Dutch' ])

```

```

table.append(['From', 'English', '', en_nl ])
table.append(['', 'Dutch', nl_en, '' ])

pytbl.make_table(table)

pytbl.set_cell_style(2,0, row_span=2, bold=True, color='gray')
pytbl.set_cell_style(0,2, column_span=2, bold=True, color='gray')

pytbl.set_cell_style(1,2, color='lightgray')
pytbl.set_cell_style(1,3, color='lightgray')
pytbl.set_cell_style(2,1, color='lightgray')
pytbl.set_cell_style(3,1, color='lightgray')

```

Out[8]:

		To	
		English	Dutch
From	English		73,431
	Dutch	3,390	

```

In [9]: data_t = pd.read_csv('unesco_xtrans_stats_years.csv', sep='\t', index_col=0)
        data_t = data_t.fillna(0)

```

6 Publications over time

The following graph indicates the number of books translated per year in Dutch and in English. Notice that because the volumes of books are significantly different, data has been plotted with two different axes.

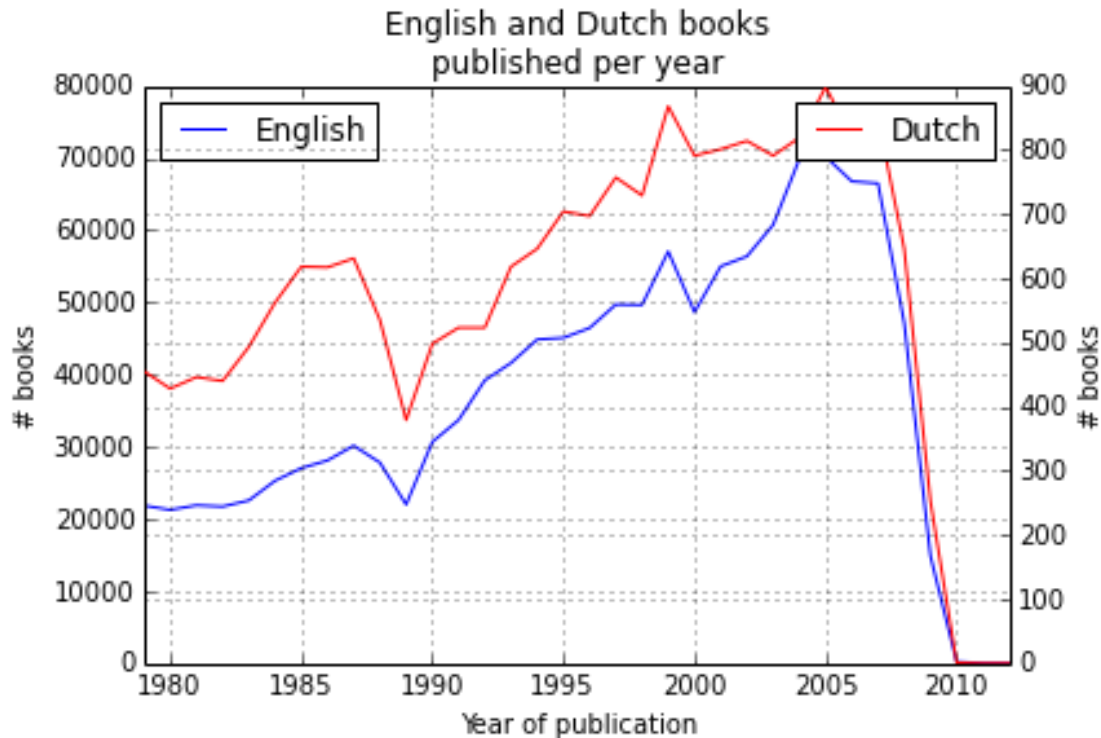
```

In [10]: ax = data_t['English'].plot(color='b', label='English')
        legend(loc='upper left')
        ax.set_ylabel('# books')
        ax.set_xlabel('Year of publication')

        twinx()
        ax = data_t['Dutch'].plot(color='r', label='Dutch')
        legend(loc='upper right')
        ax.set_ylabel('# books')

        title('English and Dutch books\npublished per year');

```



```
In [11]: data_tt = pd.read_csv('unesco_xtrans_stats_trans_years.csv', sep='\t', skiprows=1)
data_tt['Target Language'] = data_tt['Target Language'].fillna('')
del data_tt['Unnamed: 36']

In [12]: data_tt['langs'] = data_tt.apply(lambda x: x['Original language'] + ' - ' + x['Target Language'])
del data_tt['Original language']
del data_tt['Target Language']
data_tt = data_tt.fillna(0)
data_tt = data_tt.set_index('langs')

data_tt = data_tt.transpose()
```

7 Translations over time

The following graph indicates the number of books which have been translated from Dutch to English and from English to Dutch over the years. Again, notice that the plot has been drawn with two axes.

```
In [13]: ax = data_tt['English - Dutch'].plot(color='b', label='EN -> NL')
legend(loc='upper left')
ax.set_ylabel('# books')
ax.set_xlabel('Year of publication')

twinx()
ax = data_tt['Dutch - English'].plot(color='r', label='NL -> EN')
legend()
ax.set_ylabel('# books')
```

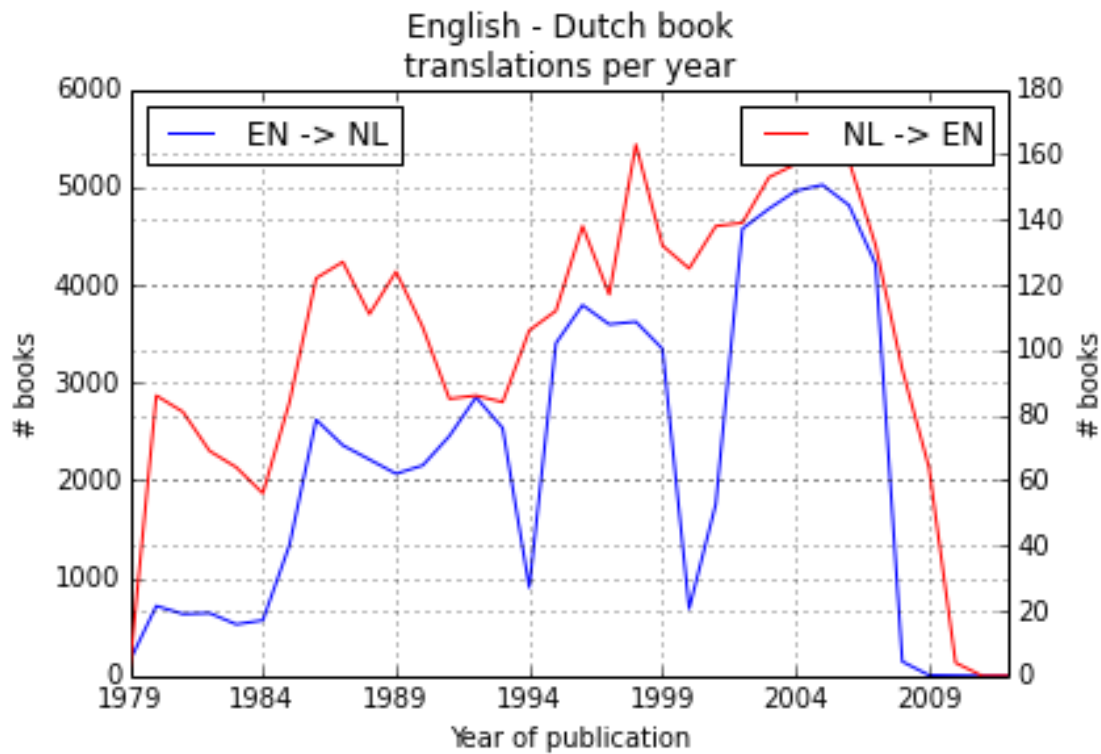


```

legend(loc='upper right')

title('English - Dutch book \ntranslations per year');

```

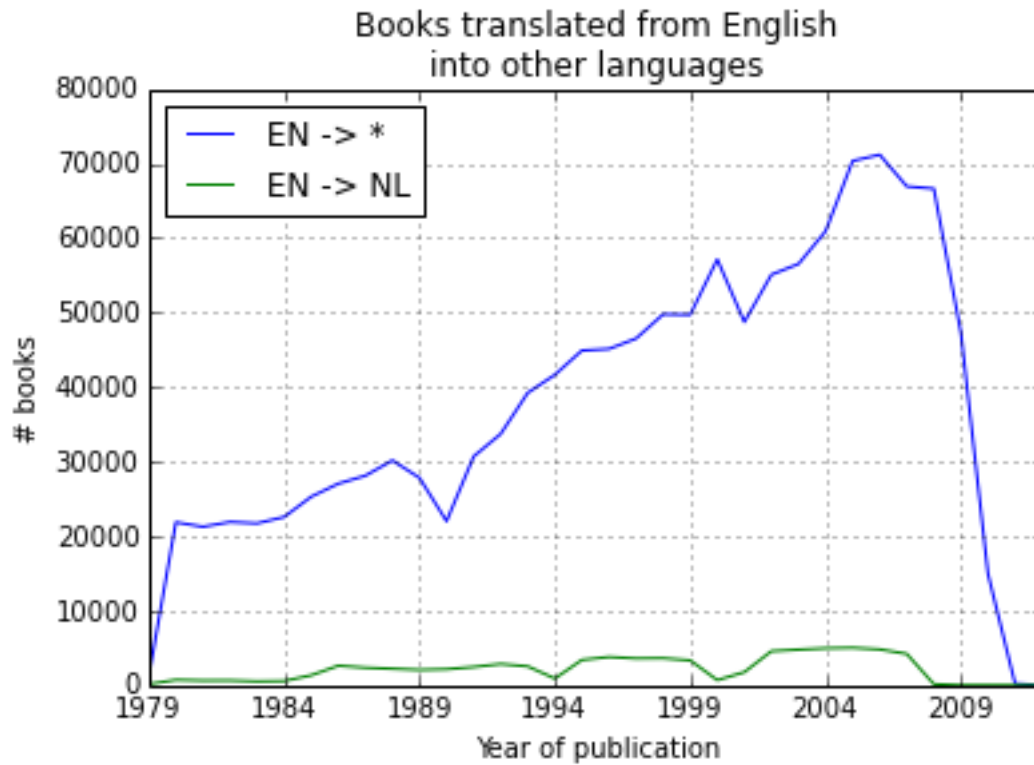


The following graphs illustrate the volume of books that have been translated over the years into and from English / Dutch.

```

In [169]: from_lang = 'English'
          to_lang = 'Dutch'
          cols = [ col for col in data_tt.columns if col.startswith(from_lang + ' - ') ]
          data_tt[cols].transpose().sum().plot(label='EN -> *')
          data_tt[from_lang + ' - ' + to_lang].plot(label='EN -> NL')
          legend(loc=0)
          xlabel('Year of publication')
          ylabel('# books')
          title('Books translated from English \ninto other languages');

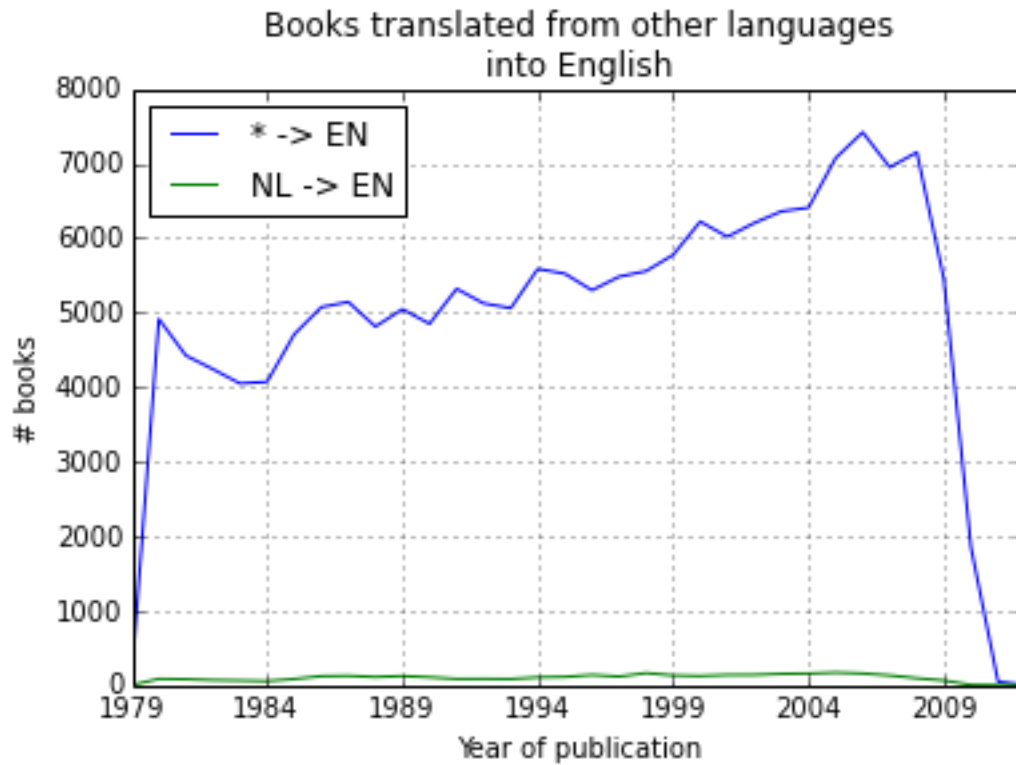
```



```
In [170]: from_lang = 'Dutch'
to_lang = 'English'
cols = [ col for col in data_tt.columns if col.startswith(from_lang + ' - ')]
data_tt[cols].transpose().sum().plot(label='NL -> *')
data_tt[from_lang + ' - ' + to_lang].plot(label='NL -> EN')
legend(loc=0)
xlabel('Year of publication')
ylabel('# books')
title('Books translated from Dutch \ninto other languages');
```



```
In [171]: from_lang = 'Dutch'
to_lang = 'English'
cols = [ col for col in data_tt.columns if col.endswith(' - ' + to_lang)]
data_tt[cols].transpose().sum().plot(label='* -> EN')
data_tt[from_lang + ' - ' + to_lang].plot(label='NL -> EN')
legend(loc=0)
xlabel('Year of publication')
ylabel('# books')
title('Books translated from other languages \ninto English');
```



```
In [172]: from_lang = 'English'
to_lang = 'Dutch'
cols = [ col for col in data_tt.columns if col.endswith(' - ' + to_lang)]
data_tt[cols].transpose().sum().plot(label='* -> NL')
data_tt[from_lang + ' - ' + to_lang].plot(label='EN -> NL')
legend(loc=0)
xlabel('Year of publication')
ylabel('# books')
title('Books translated from other languages \ninto Dutch');
```

