



Alignment Merging

Jean Mark Gawron

Linguistics 581

San Diego State University

gawron@mail.sdsu.edu

<http://www.rohan.sdsu.edu/~gawron>

Alignment merging

```
GROW-DIAG-FINAL(e2f, f2e):
```

```
  neighboring = ((-1,0), (0,-1), (1,0), (0,1), (-1,-1), (-1,1), (1,-1), (1,1))
```

```
  alignment = intersect(e2f, f2e);
```

```
  GROW-DIAG(); FINAL(e2f); FINAL(f2e);
```

```
GROW-DIAG(): iterate until no new points added
```

```
  for english word e = 0 ... en
```

```
  for foreign word f = 0 ... fn
```

```
    if ( e aligned with f )
```

```
      for each neighboring point ( e-new, f-new ):
```

```
        if ( ( e-new not aligned and f-new not aligned ) and
```

```
            ( e-new, f-new ) in union( e2f, f2e ) )
```

```
          add alignment point ( e-new, f-new )
```

```
FINAL(a):
```

```
  for english word e-new = 0 ... en
```

```
  for foreign word f-new = 0 ... fn
```

```
    if ( ( e-new not aligned or f-new not aligned ) and
```

```
        ( e-new, f-new ) in alignment a )
```

```
      add alignment point ( e-new, f-new )
```

Your assignment Nov 6

1. Implement GROW-DIAG-FINAL merging.
 - (a) Train an f2e IBM 2 model of the 1000 and 10000 sentence corpora (you have already done this), compute Viterbi alignments and save them to, say, `short_aligned.0.1001.fr-en.fr.aligned`
 - (b) Train an e2f IBM 2 model of the 1000 and 10000 sentence corpora (you havent done this), compute Viterbi alignments and save them to, say, `short_aligned.0.1001.fr-en.en.aligned`

This can be done just by swapping the arguments `e_file` and `f_file` in your top level trauning function. However, be sure that the code that outputs the Viterbi alignments to the `en_fr` file swaps the source and target numbers:
`0-3→3-0`

so that the alignment pairs in the `en_fr` and `fr_en` files are comparable.
 - (c) Write code implementing the GROW-DIAG-FINAL algorithm, outputting the merged alignments to, say, `short_aligned.0.1001.fr-en.gdf.aligned`
2. Turn in your entire new IBM 2 new system with merging, and your scores with and without merging.