Grote verzamelingen van vertaalde teksten – zogenaamde *parallelle corpora* - worden vaak automatisch op zins- en woordniveau gealigneerd om automatische vertaalsystemen op te trainen. Soms voegt men ook automatisch syntactische bomen aan de zinnen toe om meer taalkundige informatie eruit te kunnen halen. Als die bomen aan beide kanten verschijnen en de boomknopen ook worden gealigneerd, is er sprake van een *parallelle treebank*.

De beste vertaalsystemen zijn bijna of helemaal puur statistisch, maar in recente jaren ontstond er een grotere nadruk op de integratie van meer taalkundig gemotiveerde data, waaronder ook het gebruik van parallel treebanks. Ze zijn echter alleen op een zeer grote schaal bruikbaar, omdat er door zo een systeem veel te leren is van hoe een taal typisch naar een andere moet worden omgezet. Daarom onderzoeken we technieken om automatisch de boomknopen accuraat te aligneren. Een bijkomend motief is het feit dat parallel treebanks ook voor andere applicaties bruikbaar zijn en als taalbronnen zelf van wetenschappelijk belang zijn. Het hele proces van het aligneren van knopen noemen wij *tree alignment.*

Wij vinden dat een combinatie van statistiche en regelgebaseerde technieken met relatief weinig trainingsgegevens en weinig features zeer accurate alignments kan produceren. Ten slotte vinden we dat, wanneer wij alignments die relatief heel veel knopen aligneren – al zijn sommigen soms fout – op een syntactisch gebaseerde systeem toepassen, dat tot verbeterde automatische vertaling leidt, in vergelijking met hetzelfde systeem die op minder maar meer accurate alignments getrained is.