1. ***Phrase Structure Trees***

EN

A small town with two minarets glides by .

(Chunks (Str A) (Str small) (Str town) (Prep with) (Str two) (Str minarets) (Str glides) (Str by) (Punct '.'))

Shenzhen's traffic police have opted for unconventional penalties before . .

(Chunks (Str Shenzhen's) (Str traffic) (Str police) (have have) (Str opted) (Str for) (Str unconventional) (Str penalties) (Adv before) (Punct '.') (Punct '.'))

United States troops now carry atropine and autoinjectors in their first-aid kits to use in case of organophosphate nerve agent poisoning

(Chunks (Str United) (Str States) (Str troops) (Adv now) (Str carry) (Str atropine) (Conj and) (Str autoinjectors) (Prep in) (NP_poss (Pron_poss their)) (Str 'first-aid') (Str kits) (Prep to) (Str use) (Prep in) (Str case) (Prep of) (Str organophosphate) (Str nerve) (Str agent) (Str poisoning))

IS

Lítill bær með tvo bænaturna líður hjá .

(Chunks (Str Lítill) (Str bær) (Str með) (Str tvo) (Str bænaturna) (Str líður) (Str hjá) (Punct '.'))

Umferðarlögreglan í Shenzen hefur áður gripið til óhefðbundinna refsinga .

(Chunks (Str Umferðarlögreglan) (Str í) (Str Shenzen) (Str hefur) (Str áður) (Str gripið) (Str til) (Str óhefðbundinna) (Str refsinga) (Punct '.'))

Bandarískir hermenn bera nú atrópín og sjálfsdælingartæki í sjúkratöskum sínum til að nota ef taugaeitrun á lífrænum fosfat verður

(Chunks (Str Bandarískir) (Str hermenn) (Str bera) (Str nú) (Str atrópín) (Str og) (Str sjálfsdælingartæki) (Str í) (Str sjúkratöskum) (Str sínum) (Str til) (Str að) (Str nota) (Str ef) (Str taugaeitrun) (Str á) (Str lífrænum) (Str fosfat) (Str verður))

2. ***Testing the Grammar on Last Week's Corpus***

English Corpus:

UDScore {udScore = 0.5327957396139215, udMatching = 22, udTotalLength = 297, udSamesLength = 160, udPerfectMatch = 2}

It can be noticed that the overall score (udScore) has the value of 0.5328, which would suggest that around 53.28% of the tokens were correctly parsed by using English.dbnf grammar. While parsing more than half the tokens properly, it is clear that there is still room for improvement.

It can also be mentioned that there are 2 sentences which have been parsed completely identically by using the grammar.

Icelandic Corpus:

UDScore {udScore = 0.4204200116301341, udMatching = 23, udTotalLength = 289, udSamesLength = 115, udPerfectMatch = 1}

The overall score (udScore) has the value of 0.4204, which would suggest that around 42.04% of the tokens were correctly parsed by using English.dbnf grammar. It is clear that less than half of the tokens are parsed correctly, which is understandable, as there is a significant number of differences between the English and the Icelandic grammar.

It can also be mentioned that there is 1 sentence which has been parsed completely identically by using the grammar.

3. Adjusting the Grammar According to the Icelandic Grammar

The grammar was modified largely by changing its lexicons to better suit the Icelandic language, but also through changes in its sentences, indirect questions, adverbials, coordination and noun phrases. The changes aren't comprehensive, partially due to Icelandic's similarities to English, but also due to my lack of in-depth knowledge of Icelandic grammar and special cases.

The results when using the new grammar are the following:

Icelandic

UDScore {udScore = 0.4440256330618425, udMatching = 23, udTotalLength = 289, udSamesLength = 119, udPerfectMatch = 1}

The overall score for Icelandic is now 0.444, which means that 44.4% of the tokens are parsed correctly, a 2.4% increase of the previous grammar. While still having a noticeable amount of errors, this grammar is a clear improvement compared to the English one in regards to the parsing of Icelandic language.

Italian

UDScore {udScore = 0.23946632979476598, udMatching = 169, udTotalLength = 4790, udSamesLength = 927, udPerfectMatch = 3}

The overall score of Italian is 0.2394, thus amounting for a correct parsing of just 23.94% of the tokens, this is to be expected as Italian is part of a completely different branch of Indo-European languages.

Finnish

UDScore {udScore = 0.27613186813186813, udMatching = 5, udTotalLength = 97, udSamesLength = 23, udPerfectMatch = 0}

The overall score of Finnish is 0.2761, which makes sense, as despite not being an Indo-European language, Finnish has been greatly influenced by North Germanic languages throughout its existence.

Czech

UDScore {udScore = 0.34999233373586847, udMatching = 41, udTotalLength = 773, udSamesLength = 255, udPerfectMatch = 0}

The overall score of Czech is 0.3499, thus 34.99% of the tokens being parsed correctly. This is a surprising result, which could be explained by the potential similarities between Czech and other Germanic languages, due to their geographic proximity but also due to their cultural (and linguistic) exchanges.