

構文厳密な日本語話者のための暗号言語 (仮称:Cyphanese)

ゐてるま

2020 年 5 月 24 日

概要

本論考では構文厳密性を保ち、日本語話者が容易に習得可能でありながら、暗号を直感的に扱えるセキュアな暗号言語の外観を示す。

また、言語の暗号化に用いられるアルゴリズムや、構文解析器等は Haskell によって実装される。高校数学程度の数学的知識、日本語文法、および様相述語論理等の知識を要求するが臨機に説明を加える。

1 Cyphanese の特徴と概観

Cyphanese は構文厳密性を保つ暗号言語であり、日本語話者の学習容易さを一つの指標として掲げている。目的の一つは日本語話者間のセキュアかつ厳密な情報伝達を可能にし、かつ容易にそれを用いてコミュニケーションを図るためである。また、Cyphanese は特定の一言語を指す名称ではなく、この文法システムと理論を用いて生成された言語の総称である。この言語の話者は相手と事前に交換した鍵を用いることで言語を生成し、セキュアな情報伝達とコミュニケーションを図れる。

この言語は日本語の単語を基盤とした a posteriori な言語であるが、暗号化を施すため都度単語の語形が変化することがある。そのため、すきえんていあ氏 (@cicada3301__kig) が開発した自然言語と人工言語の区別の指標となりうるユーフォニー指数を使用して暗号化アルゴリズムを作成することで、暗号言語でありながら自然言語のように容易に発話可能という特徴を持っている。

ユーフォニー指数 (Euphony index;)

$$E(\alpha, \beta, \gamma, \delta, \varepsilon) = \begin{cases} E_0 & (\varepsilon \neq 2) \\ -0.04E_0^2 + 1.4E_0 & (\varepsilon = 2) \end{cases}$$

$$E_0 = \frac{1}{2} \cdot \left(1 + \frac{1}{1 + \exp(0.5\alpha - 7)} \right) \cdot \left(\frac{100}{1 + \exp(-2.26\alpha - 0.0693\beta + 0.0112\gamma + 0.388\delta - 11.9)} \right)$$

α = 単語長の平均 (文字数)

β = 語頭 2 字が両方子音になる割合 (%)

γ = 語頭 2 字が両方子音で、かつ s - C - r/l/h の構造を取らない割合 (%)

δ = 文章を通じての子音字の出現割合 (%)

ε = 子音クラスをなす文字長の最頻値 (文字数)

2 構文厳密性

この言語の特徴の一つに構文厳密性がある。構文厳密性とは任意の文に対して、構文木が一意に定まることをいう。