暗号・情報保全史特論
History of Cryptograph and Signal Security Advanced Course

第2回: 古代～中世の暗号

佐藤永欣

今回の内容

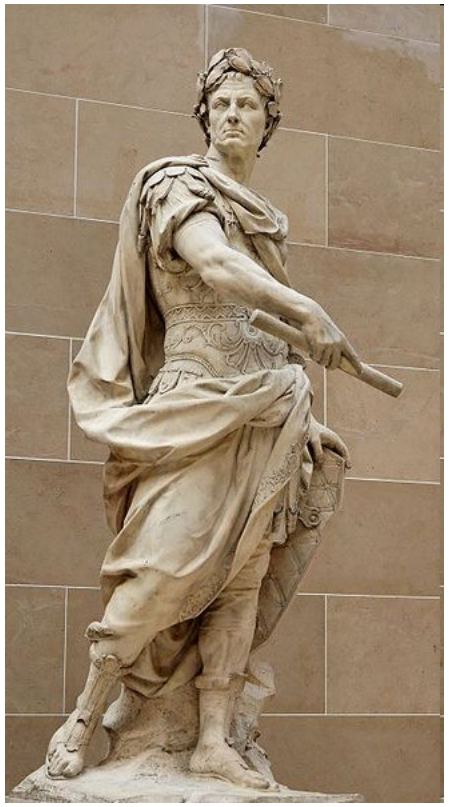
- 古代～中世ヨーロッパのcypher
- シーザー暗号

- 単一換字暗号
- ヴィジュネル暗号

- コードブックの使用

シーザー暗号

- シーザー暗号
  - ユリウス・カエサル(Gaius Julius Caesar, 英語読み:ジュリアス・シーザー, 古代ローマの将軍・皇帝・作家）が使用したと言われている
    - 「来た、見た、勝った」「賽は投げられた」
    - Julyの語源
- 暗号化の手順
  - 平文のアルファベットをn文字後ろにずらす
    - 4文字ずらす場合: a→E、b→F、c→G、…、v→Z、w→A、x→B
- 復号化の手順
  - 平文のアルファベットをn文字前にずらす
- 暗号化鍵
  - n: ずらす文字数
- 暗号化鍵は何種類?



後世のかなり美化された彫像

Wikimedia Commonsより

シーザー暗号

- いわれている/わかっていること(割とどうでもいい話)
  - 実際にカエサルがシーザー暗号を使ったかどうかは怪しい
  - ガリア戦記(カエサルが今のフランス、イギリスに攻め込んだ時の戦記。カエサルの自著)には手紙にはギリシャ語を使っていたという記述が1カ所だけある
    - ガリア人(ゲルマン人の大移動前のフランスの住民)にはラテン語を読み書きできるものがいた
    - ガリア人はギリシャ語まではさすがに習っていなかった。ローマ人のインテリには必須の教養だった
    - 暗号化云々については全く書かれていない
  - 内乱記(ポンペイウス・元老院派との内戦の記録)を含めてカエサルの著作には情報の伝達手段のことはほとんど出てこない
  - シーザー暗号に関する記述が初めて出てくるのはスエトニウスによるローマ皇帝伝
    - カエサルの時代から200年近くたって書かれた→どこまで信用できるかは不明

シーザー暗号

- どの程度有効だったのか?
  - ローマ時代は識字率が低かったことを考えると平文でも十分だった?
    - 貴族と騎士階級は読み書きできた。大商人と農場の奴隷管理者も読み書きできた
    - 書記・教師が仕事の奴隷がいた
    - 軍人: 百人隊長クラスから上は読み書きできた模様
    - ポンペイで発見された落書き: 一般市民にも読み書きができるものはそれなりにいた
  - ギリシャ語の平文を使ったという説もある(前述)
  - 単純ではあるが当時としては十分な強度がある暗号
  - ゲルマン人の大移動→西ローマ帝国滅亡→フランク王国の時期はさらに識字率が下がった。12～14世紀ぐらいまではこの状況が続く
    - 貴族でも自分の名前を書くのがやっと。修道士・司祭は読み書きできた
- どの程度の期間使われたのか?
  - 頻度分析に関する初めての記述は9世紀のアラブ
  - ヨーロッパに伝わるまで数百年?

シーザー暗号の換字表

- 換字表は比較的単純(どれか1行だけ使用する)

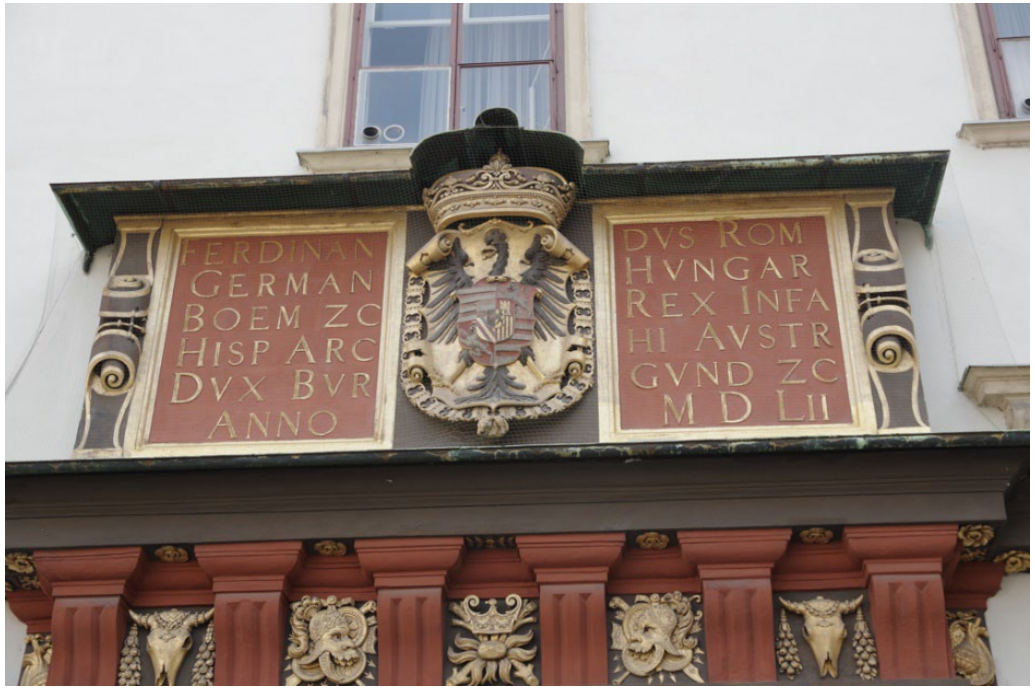
	キー																								
平文		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
暗号文	0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	1	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
	2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
	3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
	4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
	…																										
	25	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

シーザー暗号(どうでもいいおまけ)

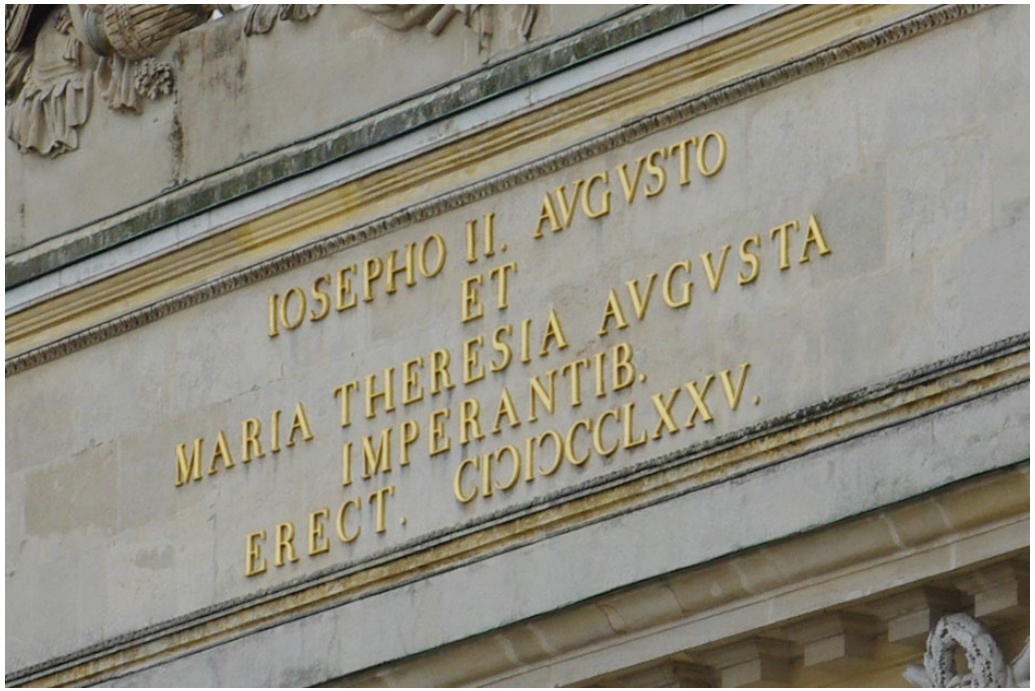
- ローマ時代(紀元前3世紀～紀元後5世紀)のアルファベットについて
  - 大文字しかなかった
    - 小文字はカロリング朝時代(9世紀ごろ)に修道院で使われ始めた
  - 碑文と筆記体でかなり字体が異なる
  - 23文字しかなかった
    - JとUはなく今の IとJ、UとVの区別はなかった。IとVは子音と母音の両方に使われていた
    - Wはなかった。VVと書かれていた
  - さらにさかのぼるとGはかなった。CがGの発音を表すために使われていた
  - Kは外来語以外には使われなかった。Kの音はCで表記した(Gの発音とは異なる)
  - AとEが続くときは合体してÆになることがある
- 算用数字はなかった
- 分かち書きはしたりしなかったり
  - 空白の代わりに碑文では･を使うことがあった
  - 単語の途中で改行しないというルールもなかった

シーザー暗号(さらにどうでもいいおまけ)

- こんな雰囲気で書かれていた(後世の碑文、どちらもウィーン市内)



Ferdinandus Rom German Hungar Boem zc Rex Infa Hisp Archi Austr Dux Burgund zc Anno MDLII (1552)
フェルディナント、ローマ、ドイツ、ハンガリー、ボヘミア等の王、スペイン王子、オーストリア大公、ブルゴーニュ等の公爵 1552



Josepo II Augusto et Maria Theresia Augusta Imperantib. Erect. MDCCLXXV (1775)　※CIↃ→M, IↃ→D
皇帝ヨーゼフ2世と女帝マリア・テレジア 1775年建築

シーザー暗号

- 解読してみよう
  - UIMFQ BDQRQOFGDMX GZUHQDEUFK
    - UNIXにはrot13コマンドがあるので力任せ攻撃も楽にできる
- どのような解読方法が考えられるか?
  - 力任せ
  - 頻度分析
  - 推測攻撃
- 解読方法から、どのような弱点が考えられるだろうか?

単一換字暗号

- シーザー暗号が知られるようになるにつれて保安度が低下
- 解読方法も知られるようになった
  - 力任せの総当たり→頻度分析
  - それだけ識字率が上がったということ?
    - 通信手段が口頭から手紙に移行すると使者を捕まえて拷問しても使者は何も知らない

暗号化用換字表

平文	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
暗号文	V	U	X	H	A	C	P	Y	G	F	R	T	K	I	E	J	W	Z	L	B	D	O	N	Q	S	M

復号化用換字表

暗号文	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
平文	e	t	f	u	o	j	i	d	n	p	m	s	z	w	v	g	x	k	y	l	b	a	q	c	h	r

単一換字暗号

- アルファベット1文字を、（別の）アルファベットに対応させ、入れ替える
  - 対応表: 換字表 → 鍵
  - 換字表の平文アルファベットと暗号文アルファベットは1対1対応の関係
  - ラテン文字→ラテン文字の対応である必要はない
    - ギリシャ文字にすることはあったらしい
    - 変体仮名も現代の日本人には暗号としては意味がありそう
  - 頻度分析が知られるようになってから、使用頻度の少ない数文字をまとめて1文字に変換する(多対1変換)ことも行われた
- 暗号鍵の数
  - 表に出てくるアルファベットの組み合わせの数だけある
  - $26 \times 25 \times \ldots \times 1 = 26! = 4.03291461 \times 10^{26}$
  - 力任せの攻撃はほぼ通じない

単一換字暗号

- 解読してみよう
  - AWM CIGNF VJXOD UXE HIPRY XSMJ AWM QKZT LXB
    - これは無理?
    - 少なくとも頻度分析を適用するには文字数が少ない
    - でも部分的に特徴があると言えばある

自然言語の書き言葉

- どのような文字言語も音声言語の影響を受ける
  - 表音文字の場合に顕著だが表意文字でも人間の言語が音声から自然に発生したものである以上、何らかの影響がある
- 子音だけの音声言語は存在しない: 必ず母音を含む音節ができる
  - 子音だけを表記する文字言語は存在する: アラビア語、ヘブライ語
    - 文法的要素により母音が自動的に決まる、母音の種類が少ない、などの理由
- このため
  - アルファベットのように母音字と子音字がわかれる文字体系では母音字(AEIOU)・半母音字(VWY)の頻度が比較的高くなる
  - かなのような表記体系の場合は出現頻度の分布がアルファベットのような場合と異なる
  - アルファベットを使用している言語でも音節単位で暗号化することもある(ルイ14世の大暗号)

自然言語の書き言葉

- 発音の都合による影響
  - 子音と母音が交互に現れる
  - 子音→子音の遷移が可能な組み合わせが限られる
    - 発音が容易・楽な組み合わせ→調音位置が大幅に動いたりはしない
    - 舌が前後に何度も往復するのは非常に発音しにくい
- 母音: 声帯が震える(有声音)
- 子音
  - 有声音: bdg(破裂音。いったん唇を閉じて急に開く)、vz(摩擦音)、mn(鼻音。唇は閉じたまま)、lr(流音)
  - 無声音(声帯は震えない): kctp(破裂音)、ΦfΘsxh(摩擦音)

自然言語の書き言葉

- 歴史的事情による綴りの影響
  - つづりはいったん固定されるとなかなか変わらない(印刷術の発明以降は特に)
  - 発音が時代で変化してもつづりは残る傾向: 日本語のカナでも実はそう
    (日本語の場合:「は」で「わ」と発音したり、は行の音が奈良時代(p)・平安時代(ph)は違ったり、最近は「つ」の母音を発音しない傾向が出たり)
- 英語の場合: 単語の起源が複数ある
  - ゲルマン語(サクソン語とか?)由来
  - フランス語由来
    - それぞれつづりに特徴がある
- qのあとほぼ必ずuはラテン語由来のローマ時代以来の特徴
  - xを使うのも同じ。ekstentではなくextentとつづる
- つづり中の文字が出現するパターンができる
  →頻度分析以前の単一換字暗号でも解読の糸口になった

おまけ: 単一換字暗号のバリエーション

- 暗号化するときの変換先が平文と同じ文字体系である必要はない

i am hungry

- Arthur Conan Doyle, “The Dancing Men”, *The Return of Sherlock Holmes*
  - シャーロック・ホームズシリーズの短編
  - 踊る人形などの題名で和訳が数種類ある(青空文庫に2編収録)
  - 頻度分析による暗号解読の手順が謎ときの重要な手段になっている
    - https://www.aozora.gr.jp/cards/000009/card50713.html

ヴィジュネル暗号

- 単一換字暗号の解読法が知られるようになったため開発された(16世紀中ごろ)
  - 頻度分析がヨーロッパじゅうに広まった
- 多表式暗号(一応): 文字の出現頻度の特徴を消すことが目的
  - ビジュネル暗号では換字表は広く知られている
    - 実はシーザー暗号を少々複雑にしたもの
- 暗号化手順
  - 暗号鍵の単語を適当に用意 → 1文字ずつ使用
    - 暗号鍵が KEY のとき、K、E、Y、K、E、Yの順に繰り返して、1文字ずつ鍵として使用
  - ビジュネル暗号表から平文の文字の行を取り出す
  - 鍵の文字の列を取り出す
  - 行と列の交点が暗号化された文字
- 300年ぐらいは解読困難なままだった(ヨーロッパの暗号関係者の主な関心事がコードの利用だったため)

ヴィジュネル暗号の暗号表

鍵: KEY
平文　: hello world
暗号文: RIJVS UYVJN

鍵の文字

平文の文字	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
b	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
c	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
d	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
e	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
f	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
g	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
h	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
i	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
j	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
k	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
l	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	
…	…	…																								

(˘ω˘).。oO(見ての通りシーザー暗号に毛が生えただけ。暗号の文字のABC..を逆順にしてZYX...にしたものもある)

ヴィジュネル暗号

- 数学的には
  - A~Zを0~25と考える
  - 平文の文字をT、暗号文の文字をC、鍵の文字をKとする
  - 暗号化: $C \equiv T+K \pmod{26}$
  - 復号化: $T \equiv C-K \pmod{26}$
- 弱点
  - 鍵が1文字だけならシーザー暗号と変わらない（ずらしてるだけ）
  - 鍵に使われる文字が周期的に変わる
    - 鍵の長さをLとすると、暗号文のL文字ごとに同じ変換をされた文字が現れる
    - つまり、L文字ごとに同じ平文が偶然存在すると????
  - 鍵の文字数さえわかれば攻撃は簡単
    - といっても、単一換字暗号よりも多くの暗号文が必要
  - 鍵はたいていは適当な単語 → 長くできない → 鍵長20文字ぐらいまでの範囲でいろいろ試せばいい
- 元々のヴィジュネルの表はアルファベットがランダムな表だったが、シーザー暗号式にアルファベット順に並んでいるものが使われた
- ヴィジュネル暗号の前段階としてトリテミウス多表があった; 暗号鍵がなく、多表の各行を上から順に使っていた
  - それでも平文の頻度の特徴は消せる。が、暗号強度的には弱い

多表式暗号の展開

- ヴィジュネル暗号は16世紀から19世紀後半まで、「解読不能の暗号」と言われて使われ続けた
  - 手作業で暗号化・復号化するためあまり複雑な暗号も使えなかった
- ヴィジュネル暗号でも安全とは言えない
  - カシスキーの方法による鍵長推定
  - クリブが表れやすい: 暗号鍵”KEY”の例
  - government of the people by the people for the people
  - QSTOVLWILD SD DLC ZIMZPC LC RRI NOSNVI DYV RRI NOSNVI
  - 鍵長の整数倍の位置に同じ平文があれば同じ暗号文になる
    - *偽のクリブ*も出る
    - 十分長い鍵を使うことで対抗はしていた

多表式暗号の展開

- 鍵長について
  - 長ければ長いほどいい
  - 詩や本の文章などが鍵として使われる時代が長かった
    - とはいえ常識からかけ離れた語句を鍵にすると強度が上がることは認識されていた
    - 意味のある語句を鍵に使うと平文に入っていそうな単語から鍵を推測されるのも認識されていた
    - ランダムなキーを使うようになったのは20世紀に入ってから(!!!!!)
- 鍵長を事実上無限にする工夫: ヴィジュネルによる自動鍵
  - 鍵の先頭だけ決めておく(下の例では1文字だがもっと長くてもいい)
  - 平文か暗号文を次の鍵の文字にする

原文:　ars longa vita brevis
鍵:　　*D*AR SLONG AVIT ABREVI
暗号文: DRJ DZBTG VDBT BSVZDA

  - 暗号化・復号化の過程で1文字でもミスするとその文字以降が崩壊する

コードブックの使用

- 平文の単語を別の単語に置き換えるタイプ
  - 主に手紙の時代: 手紙を盗まれなければ大丈夫。ステガノグラフィの併用もあった
  - 外交用暗号での使用が長かった
  - ヴィジュネル暗号・単一換字暗号との併用もあったがコードブック単独での使用も多かった
  - コードブックにない語はそのまま
- 平文の単語を数桁の数字・数文字のアルファベットの羅列に置き換えるもの
  - 電信と機械式暗号が普及しだしてから: 電信士etcは通信文を見る
  - すべての語をコードに置き換えるのが前提
    - コードブックにない語はつづりを分解して各文字に該当するコードに変換

転置式暗号

- 普通に文章を書く→規則に従って文字の順番を入れ替える
- Ex)
  - 右から書く
  - グルグルらせん状に書く
  - N文字ずつ区切って文章全体を並べ替える
    - 1,2,3,…,Nや1,N,2,N-1,…のような規則をあらかじめ決めておく
- 自然言語の特徴は消えない
  - 頻度分析には弱い
- 多表式よりも扱いが簡単なため「素人が使うにはちょうどいい」程度の認識で19世紀終わりごろまで使われていた