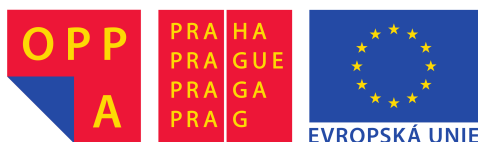

Enigma a jiné šifry

doc. RNDr. Jiří Tůma, DrSc.

Kurz vznikl v rámci projektu "Rozvoj systému vzdělávacích příležitostí pro nadané žáky a studenty v přírodních vědách a matematice s využitím online prostředí", Operační program Praha – Adaptabilita, registrační číslo CZ.2.17/3.1.00/31165.



Projekt A-NET je financován Evropským sociálním fondem, rozpočtem ČR a MHMP.

"Praha & EU: Investujeme do vaší budoucnosti."

Základní pojmy

Už od starověku se lidé snaží předávat zprávy tak, aby je mohl číst pouze adresát a nikdo jiný. Je-li zpráva napsána ručně a doručována otrokem (jak tomu bylo ve starém Řecku nebo Římě) nebo poštou dnes, tak vždy existuje nebezpečí, že se dostane do ruky někomu, komu není určena. Otrok může být zajat, pošťák se může zmýlit a doručit zprávu na nesprávnou adresu. Je-li zpráva napsána srozumitelně, přirozeným jazykem bez jakékoliv snahy skrýt její obsah, tak jí může porozumět každý, komu se dostane do ruky, pokud zná jazyk, kterým je napsána.

V novější době můžeme zprávy předávat telegrafem, radiovými vlnami, telefonem, faxem nebo e-mailem, ale nebezpečí, že budou odposlechnuty, je stále přítomné. Ve skutečnosti se od té doby nesmírně zvýšilo. Tak například radiové vysílání může slyšet každý, kdo je v dosahu vysílače a má přijímač naladěný na správnou frekvenci. Zpráva e-mailem může být doručena na spoustu nezamýšlených adres v důsledku překlepu nebo viru číhajícího v počítači.

Jakkoliv pesimisticky to může vypadat, je dobrým pravidlem předpokládat, že *každá* zpráva, která má být důvěrná, se může dostat do rukou někomu, komu není určena. Je proto projevem potřebné opatrnosti učinit kroky, které zajistí, že nezamýšlený příjemce bude mít přinejmenším velké problémy zprávu porozumět, nebo ještě lépe, vůbec ji nebude schopen přečíst.

Rozsah škody, kterou může nezamýšlené odhalení zprávy způsobit, do velké míry závisí na čase, který uplynul mezi odposlechnutím zprávy a jejím rozluštěním. V některých případech stačí, aby mezi odposlechnutím a rozluštěním zprávy uplynul jeden den nebo dokonce jenom pár hodin, aby k žádné škodě nedošlo. Například rozhodnutí akcionáře prodat nebo koupit okamžitě velký balík akcií, nebo rozkaz armádního velitele zaútočit za rozbřesku následujícího dne. V jiných případech má informace dlouhodobou hodnotu a musí být proto utajena co nejdéle. Například zprávy týkající se plánování rozsáhlých vojenských operací.

Úsilí, které musí vynaložit soupeř, oponent nebo nepřítel k tomu, aby rozluštil zachycenou zprávu, má proto velký význam. Jestliže neautorizovaný příjemce zprávy nedokáže zprávu rozluštit za použití nejlepších známých metod a nejvýkonnějších dostupných počítačů za dobu kratší, než po kterou je utajení obsahu zprávy důležité, tak si odesílatel může být relativně jistý. Nemůže si ale být *zcela* jistý, protože rozluštění nějakých dříve odeslaných zpráv může protivníkovi pomoci urychlit luštění následujících zpráv. Je také

možné, že protivník objevil nějakou metodu, kterou odesílatel nezná, a je proto schopen luštit zprávy rychleji, než si odesílatel dokáže představit. To se například přihodilo německé armádě se šifrovacím zařízením Enigma již před zahájením a později také v průběhu druhé světové války.

Ceasarova šifra

Říká se, že staří Řekové vymysleli zvláštní způsob, jak utajit zprávy před protivníkem. Zprávu napsali na oholenou hlavu otroka a pak počkali, než mu hlava znovu zaroste. Pak jej vyslali k adresátovi. Ten otrokovi hlavu zase oholil a zprávu si přečetl. To byla očividně velmi nespolehlivá a pomalá metoda. Každý, kdo ji znal a zadržel příslušného otroka, mohl zprávu snadno přečíst. Navíc odeslání zprávy trvalo celé týdny a další týdny bylo třeba čekat na odpověď.

Místo v historii kryptografie si zajistil až Julius Caesar. Ten napsal zprávu a každé písmeno zprávy nahradil písmenem, které je v abecedě o tři místa dále. Pokud by používal mezinárodní latinskou abecedu, nahradil by každé písmeno A písmenem D, písmeno B písmenem E, a tak dále, písmeno W písmenem Z, potom písmeno X písmenem A, písmeno Y písmenem B a písmeno Z písmenem C. Každé písmeno mezinárodní latinské abecedy tak nahradil písmenem, které je v následující tabulce pod ním.

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
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

Kdyby Julius Caesar chtěl někomu utajeně sdělit, že

KOSTKY JSOU VRZENY,

poslal by

NRVWNB MVRX YUCHQB.

Caesarova metoda není příliš důmyslná. Zejména proto, že na první pohled odhaluje, že zpráva se skládá ze tří slov, první a poslední jsou tvořené šesti písmeny zatímco druhé má čtyři písmena. Slabosti takového naivního systému není jednoduché odstranit, i když rozšíření abecedy z 26 symbolů na 29 nebo více tak, aby zahrnovala také interpunkční znaménka a mezeru, by učinilo rozpoznání délky jednotlivých slov *nepatrně* obtížnějším. Přesto je Caesarova šifra příkladem *šifrovacího systému* a je speciálním případem *jednoduché záměny*, šifry, kterou se budeme za chvíli zabývat.

Několik základních definic

Opakovaně budeme používat slova jako *bigram*, *kryptografie*, *zašifrovat*, a proto je nyní definujeme.

Monogram je jedno písmeno v jakémkoliv abecedě, kterou budeme používat. *Bigram* je jakákoliv dvojice *sousedních* písmen v nějakém textu. Tak například DO je bigram v českém textu, AT je bigram v anglickém textu. *Trigram* je trojice po sobě následujících písmen. *Polygram* je tvořený nespécifikovaným počtem po sobě jdoucích písmen v textu. Polygram nemusí být slovem v nějakém jazyce, ale pokud se pokoušíme rozšifrovat nějakou zprávu, o které můžeme rozumně předpokládat, že je v angličtině, pak je nalezení *heptagramu* MEETING mnohem slibnější, než nalezení heptagramu QYRDBCFFK.

Symbolem rozumíme jakékoliv písmeno, číslici, interpunkční znaménko atd., které se mohou v textu vyskytnout. *Řetězec* je jakákoliv posloupnost po sobě jdoucích symbolů. *Délka* řetězce je počet symbolů, které řetězec obsahuje. Tak například A3£%\$ je řetězec délky 5.

Šifrovací systém nebo také *kryptografický systém* je jakýkoliv systém, který lze použít ke změně textu nějaké zprávy s cílem učinit ji nesrozumitelnou komukoliv jinému s výjimkou adresáta, kterému je určena.

Pokud nějaký šifrovací systém použijeme na nějakou zprávu, tak říkáme, že zprávu *šifrujeme* nebo že jsme ji *zašifrovali*.

Původní text zprávy, ještě před tím, než byl zašifrován, se nazývá *otevřený text*; poté, co byl zašifrován, se nazývá *šifrový text* nebo *šifrová zpráva*.

Opačný proces k šifrování, tj. rekonstrukce původního otevřeného textu zprávy z jeho šifrové verze, se nazývá *dešifrování* nebo *luštění*. Tato dvě slova neznamenají zcela totéž. Zamýšlený příjemce zprávy tuto zprávu dešifruje, zatímco nezamýšlený příjemce, který se snaží pochopit její obsah, ji luští, případně řeší.

Kryptografie je věda o tom, jak navrhovat a používat šifrovací systémy včetně zkoumání jejich silných stránek a slabín a jejich odolnosti vůči různým metodám útoků. *Kryptografové* jsou ti, kteří se návrhem, používáním a zkoumáním šifrovacích systémů zabývají.

Kryptoanalýza je studium metod luštění šifrovacích systémů. Ten, kdo se zabývá kryptoanalýzou, je *kryptoanalytik*, populární (a neformální) anglický výraz je *codebreaker*.

Kryptografie a kryptoanalýza jsou dvě součásti jedné vědecké disciplíny, kterou nazýváme *kryptologie*. Společnost *kryptologů* tvoří jak kryptografové tak kryptoanalytici.

Kryptografové a kryptoanalytici jsou soupeři, každý z nich se snaží přelstít toho druhého. Představuje si sám sebe v postavení toho druhého a klade si otázky jako “Kdybych byl na jeho místě, co bych asi udělal, abych mě porazil?”. Obě strany, které se pravděpodobně nikdy nesetkají, jsou zapojené do fascinujícího intelektuálního souboje a v sázce může být opravdu hodně.

Tři fáze luštění: identifikace, prolomení a nastavení

Když kryptoanalytik poprvé vidí zašifrovanou zprávu, jeho první otázkou je odhalit, jaký šifrovací systém byl použit. Může to být nějaký už známý šifrovací systém nebo zcela nový. Tomu se říká problém *identifikace*. K tomu vezme v úvahu veškeré dostupné současné informace. Například jaký typ šifrovacího systému odesílatel, je-li znám, dosud používal, nebo jaké nové šifrovací systémy se v nedávné době objevily. Potom začne zkoumat úvod zprávy. Ten může obsahovat informace, které mají zamýšlenému příjemci zprávy pomoci v jejím dešifrování, ale mohou být také užitečné kryptoanalytikovi. Potom analyzuje celou zprávu. Je-li zpráva příliš krátká, tak může být nemožné pokročit dále a kryptoanalytik musí čekat na další zprávy. Je-li zachycená zpráva dostatečně dlouhá nebo pokud má kryptoanalytik k dispozici několik zpráv, tak může použít řadu matematických (zejména statistických) testů, které mu s jistotou odhalí, jestli byla použita například kódová kniha nebo nějaký jednoduchý šifrovací systém případně něco dokonalejšího.

Pokud se kryptoanalytikovi podařilo identifikovat použitý šifrovací systém, může odhadnout, kolik šifrových zpráv bude potřebovat, aby měl přijatelnou naději, že jej *prolomí*, tj. že přijde na to, jakým způsobem systém zprávy šifruje. Jde-li o jednoduchý systém, u kterého nedochází k podstatným změnám mezi jednotlivými zprávami, jako je například kódová kniha, jednoduchá záměna nebo transpozice, tak může být schopen zprávy rozluštit bez velkých problémů. Daleko pravděpodobnější ale je, že některé části šifrovacího systému se mění zprávu od zprávy a kryptoanalytik potřebuje napřed určit, jaké části systému se nemění. Tak například přístroj pro německý šifrovací systém Enigma používaný ve druhé světové válce obsahoval několik rotorů. Uvnitř těchto rotorů bylo propojení pomocí drátů. Propojení uvnitř rotorů se neměnilo, ale jejich pořadí a vzájemné pootočení se měnilo každý den. Propojení uvnitř jednotlivých rotorů tak tvořilo pevnou (neměnnou) část systému, zatímco jejich pořadí bylo proměnlivé. Prolomení je nejobtížnější fází luštění šifrových zpráv. Může trvat týdny a měsíce a vyžadovat použití pokročilých matematických metod, využití chyb obsluhy systému nebo informací získaných špionáží.

Pokud se už podařilo určit všechny pevné součásti šifrovacího systému, je třeba ještě určit, jak se mění proměnlivé části. Například počáteční nastavení rotorů v přístroji Enigma, které se měnilo s každou zprávou. Tomu se říká problém *nastavení*. Pokud se podaří vyřešit i tento problém, je možné zprávy rozluštit.

Prolomení se tak týká celého šifrovacího systému, zatímco *nastavení* se

vztahuje k luštění jednotlivých zpráv.

Šifry a kódy

Budeme rozlišovat mezi *šiframi* a *kódy*. Pomocí šifry nebo obecněji šifrovacího systému se odesílatel a adresát snaží *utajit* obsah zprávy před nepovolanou osobou. Smyslem *kódu* naopak není zprávu utajit, ale upravit ji tak, aby ji bylo možné přenést nějakým kanálem. Příkladem kódu je třeba Morseova abeceda. Morseova abeceda používá tři symboly — tečku, čárku a mezeru — k úpravě textu napsaného v běžném jazyce tak, aby jej bylo možné přenést pomocí nemodulovaného radiového signálu. Rovněž ASCII kód nahrazuje jednotlivá písmena a další symboly pomocí posloupností osmi bitů, které jsou uzpůsobené pro přenos informace pomocí elektrického proudu.

Šifrovací systémy vytváří šifrovou zprávu z otevřeného textu pomocí nějakého pravidla – *algoritmu*. Až do nástupu počítačů v kryptografii dominovaly tři základní metody a mnohé šifrovací systémy byly založené na jedné z nich, případně na nějaké kombinaci dvou nebo všech třech metod. První idea spočívá v zamíchání písmen abecedy stejně jako se míchá balíček karet. Tuto metodu nazýváme *záměna* nebo také *substituce*. Můžeme buď použít jedno a totéž zamíchání pro každé písmeno celého otevřeného textu, v tom případě mluvíme o *jednoduché záměně (substituci)*. Nebo můžeme pro každé písmeno otevřeného textu abecedu nově zamíchat, takovou šifru nazýváme *polyalfabetická substituce*. Polyalfabetickou substituci můžeme realizovat například tak, že nahradíme písmena v otevřeném textu zprávy čísly podle nějakého pravidla, např. A=0, B=1, C=2, . . . , Z=25. K nim potom pořadě přičítáme nějaká jiná čísla, kterými rovněž mohou být písmena nahrazená čísly, říká se jim *klíč*. Pokud je součet větší než 25, odečteme od něho 26, abychom dostali opět nějaké přirozené číslo mezi 0 a 25. Tomu se říká *modulární aritmetika* s modulem 26. Výsledná čísla potom zpětně nahradíme písmeny. Pokud jsou přičítaná čísla klíče vytvořena nějakým dostatečně nepředvídatelným způsobem, pak je velmi obtížné, pokud ne zcela nemožné, šifrovou zprávu bez znalosti klíče rozluštit.

Druhá idea spočívá v zamíchání *pořadí* písmen v otevřeném textu. Těmto šifram říkáme obecně *transpoziciční šifry* nebo krátce *transpozice*. Nejjednodušším příkladem transpozic jsou přesmyčky nebo lištovky ze zábavných koutků nedělních novinových příloh.

Třetí idea spočívá v šifrování pomocí *kódové knihy*. Kódová kniha je vlastně slovníkem, ve kterém jsou běžné fráze tvořené jedním nebo více písmeny, čísly nebo slovy, typicky nahrazované čtveřicemi nebo peticemi písmen nebo čísel. Říká se jim *kódové skupiny*. V případě nejčastěji používaných výrazů nebo písmen může kódová kniha obsahovat několik *kódových skupin* s

cílem umožnit odesílateli výběr. Záměrem je učinit identifikaci nejužívanějších frází obtížnější. Tak například ve čtyřmístném kódu může být pro slovo “pondělí” několik možností: 2475 nebo 7032 nebo 6845.

Rozdíl mezi jednotlivými typy šifer je v některých případech nezřetelný. Tak například Caesarovu šifru můžeme považovat za šifru založenou na kódové knize s jednou stránkou, na které je vedle každého písmena napsáno písmeno, které je za ním v abecedě na třetím místě. Jakkoliv je Caesarova šifra jednoduchá, může být považována za ilustraci také obou předcházejících myšlenek. V prvním případě naše “míchání” abecedy spočívá v přesunutí posledních tří písmen X, Y, Z abecedy na její počátek. Ve druhém případě je klíčem prostě stále opakované číslo 3. Je to ten nejjednodušší možný klíč. V naprosté většině případů je ale rozdíl jasný. Tak například Enigma je *šifrovací* přístroj, který v žádném případě není založený na *kódové knize*.

Překlad zprávy do jiného jazyka lze považovat za šifrování pomocí kódové knihy – slovníku. To ale je význam výrazu *kódová kniha* rozšířený do krajnosti. Použití málo známého jazyka k předávání zpráv krátkodobého významu může být ale v některých případech vhodné. Tak například během druhé světové války používala americká armáda v Pacifiku k předávání telefonických zpráv menšího významu vojáky z indiánského kmene Navajů. Jejich využití bylo založené na rozumném předpokladu, že i kdyby nepřátelská japonská armáda telefonické rozhovory odposlechla, tak je velmi nepravděpodobné, že by měla k dispozici někoho, kdo by jim byl schopen rozumět.

Jiným způsobem šifrování je používání nějakého osobního těsnopisu. Tato metoda byla některými lidmi využívána už od středověku k psaní osobních deníků. Je-li k dispozici dostatek záznamů, není obtížné takovou šifru rozluštit. Pravidelné výskyty některých symbolů, jako třeba názvy dnů v týdnu, mohou poskytnout dostatečný klíč k rozluštění některých polygramů. Daleko hlubším příkladem je rozluštění starodávných Mykénských textů založených na symbolech nahrazujících řeckou abecedu Michaelem Ventrisem v roce 1952. Mnohem známější je rozluštění hieroglyfů francouzským badatelem J.F. Champollionem, ten měl ale narozdíl od Ventrise k dispozici “paralelní text” ve známém jazyce. Na rozluštění čekají ještě další starodávné texty.

Nástup počítačů a možnost vytváření složitých elektrických obvodů na silikonových čípech zcela proměnily jak kryptografii tak kryptoanalýzu. V důsledku toho jsou některé z nedávných šifrovacích systémů založené na pokročilých matematických metodách, které vyžadují výkonné výpočetní a elektronické prostředky, a byly tak prakticky nepoužitelné v předpočítačové době.

Posuzování spolehlivosti šifrovacích systémů

Pokud někdo navrhuje nový šifrovací systém, tak je důležité posoudit jeho sílu — odolnost vůči všem známým útokům za předpokladu, že kryptoanalytik zná typ tohoto šifrovacího systému, ale nezná všechny jeho detaily. Odolnost šifrovacího systému může být posuzována ve třech různých situacích:

1. kryptoanalytik má k dispozici pouze šifrové texty,
2. kryptoanalytik má k dispozici jak šifrové texty tak také jim odpovídající otevřené texty,
3. kryptoanalytik má k dispozici šifrové texty k otevřeným textům, které si sám zvolil.

První situace je ta “obvyklá”. Šifrovací systém, který lze v této situaci rozluštit za rozumnou dobu, by neměl být používán. Druhá situace může nastat, pokud je identická zpráva zašifrována jak pomocí nového systému, tak také pomocí “starého” systému, který už kryptoanalytik umí luštit. Takové situace, které jsou významným porušením pravidel bezpečnosti předávání zpráv, nezřídka nastávají. Třetí situace může nastat pokud kryptograf chce posoudit sílu navrženého šifrovacího systému. Vyzve kolegy, aby jeho systém rozluštili, a dovolí jim nadiktovat si zprávu, která má být zašifrována. Toto je standardní metoda posuzování nových šifrovacích systémů. Pro kryptoanalytika je velmi zajímavý problém jak formulovat texty tak, aby mu po zašifrování poskytly co nejvíce informací o detailech systému. Podoba těchto zpráv závisí na tom, jak šifrování probíhá. Druhá a třetí situace může také nastat, pokud má kryptoanalytik k dispozici špiona v protivné kryptografické organizaci. Šifrovací systém, který nelze rozluštit ani za třetí situace, je opravdu silný. Po takových kryptografové touží a kryptoanalytici se jich obávají.

Kódy, které odhalují a opravují chyby

Zcela jiná skupina kódů je navržena s cílem zajistit *přesnost* přenášené informace, nikoliv skrýt její *obsah*. O takových kódech říkáme, že *odhalují chyby*, případně že *opravují chyby*. Tyto kódy jsou předmětem intenzivního matematického výzkumu. Jsou používány od nejranějších počátků počítačů, aby je ochránily před chybami v operační paměti nebo v datech uložených na magnetické pásce. První kódy, jako jsou třeba Hammingovy kódy, dokážou správně opravit jednu chybu v posloupnosti sedmi bitů. Novější verze

používané například při přenášení informací z Marsu získaných sondou Mariner dokážou správně opravit až sedm chyb v každém vysílaném 32-bitovém slově. Chrání tak přenášenou informaci před značným poškozením, ke kterému může dojít během dlouhé cesty z Marsu na Zemi.

Příkladem kódu jiného druhu, který dokáže jednu chybu odhalit ale ne opravit, je mezinárodní číselný systém knih ISBN (International Standard Book Number). Každé číslo ISBN $a_1a_2a_3 \cdots a_{10}$ je tvořené buď posloupností deseti cifer nebo posloupností devíti cifer následovaných písmenem X (které je interpretováno jako číslo 10), a dává možnost překontrolovat, zdali číslo neobsahuje jednu chybu. Kontrolu provedeme tak, že spočítáme *kontrolní součet*

$$1a_1 + 2a_2 + \cdots + 10a_{10} = \sum_{i=1}^{10} ia_i.$$

Cifry jsou většinou rozdělené do čtyř skupin oddělených pro pohodlí pomlčkou nebo mezerou. První skupina označuje jazykovou oblast, druhá nakladatelství, třetí je sériové číslo knihy v nakladatelství a poslední skupina je *kontrolní cifra*.

Kontrolní součet musí být dělitelný číslem 11. Pokud není, tak je v čísle ISBN chyba. Tak například některé části tohoto textu jsou inspirovány knihou, která má ve vázané verzi jako ISBN číslo 0 521 81054 X. Kontrolní suma se pak rovná

$$1 \cdot 0 + 2 \cdot 5 + 3 \cdot 2 + 4 \cdot 1 + 5 \cdot 8 + 6 \cdot 1 + 7 \cdot 0 + 8 \cdot 5 + 9 \cdot 4 + 10 \cdot 10 = 242 = 22 \cdot 11.$$

Naproti tomu 0 987 65432 1 má kontrolní součet

$$1 \cdot 0 + 2 \cdot 9 + 3 \cdot 8 + 4 \cdot 7 + 5 \cdot 6 + 6 \cdot 5 + 7 \cdot 4 + 8 \cdot 3 + 9 \cdot 2 + 10 \cdot 1 = 19 \cdot 11 + 1,$$

a musí proto obsahovat aspoň jednu chybu.

Kód ISBN dokáže *odhalit* jednu chybu, neumí ji ale *opravit*. Pokud jsou v čísle ISBN aspoň dvě chyby, může kontrolní součet dokonce naznačit, že jde o správné číslo, i když tomu tak není.

Oblast kódů odhalujících a opravujících chyby je založena na dosti pokročilé matematice a v tomto textu se jí nebudeme zabývat.

Jiné metody skrývání zpráv

Existují jiné metody skrývání smyslu nebo obsahu zpráv, které nejsou založené na šifrách. První dvě nejsou z našeho pohledu relevantní, zaslouží si ale zmínku. Jsou to

1. použití “neviditelných” inkoustů,

2. využití *mikroteček*, drobounkých fotografií zprávy na mikrofilmu přilepených na zprávu na předem domluveném místě,
3. “vlození” zprávy dovnitř jinak nevinně vyhlížejícího textu, *slova* nebo *písmena* zprávy jsou rozptýlena podle nějakého pravidla v textu, který sám o sobě není utajený.

První dvě metody jsou s různým úspěchem využívány mnohými špiony. Třetí metoda může být rovněž využívána vězni nebo válečnými zajatci v dopisech domů k předávání informací o umístění táborů a podmínkách v nich. Touto metodou se zabývá obor nazývaný *steganografie*.

Příklady v této a následujících přednáškách budou založené téměř výhradně na mezinárodní abecedě tvořené 26 písmeny, žádná diakritika nebude používána. V některých případech budeme používat rozšířenou abecedu, která obsahuje také mezeru a interpunkční znaménka jako je tečka, čárka, dvojtečka, atd.

Teoreticky není obtížné modifikovat příklady tak, aby obsahovaly také více písmen používaných v jiných jazycích. Pokud je ale k šifrování zpráv používáno nějaké šifrovací zařízení, může požadavek zvětšit počet písmen abecedy vést k nutnosti změnit konstrukci zařízení.

Modulární aritmetika

V kryptologii je často třeba sčítat posloupnost (proud) čísel s jinou posloupností (proudem) čísel nebo odečítat jednu číselnou posloupnost od jiné. Zpravidla nejde o obvyklé sčítání nebo odčítání celých čísel, ale o *modulární aritmetiku*. V modulární aritmetice sčítáme, odčítáme a násobíme čísla vzhledem k nějakému předem zvolenému číslu, kterému říkáme *modul*. Typické příklady modulů jsou 26, 10 nebo 2. Ať používáme jakýkoliv modul, je každé číslo ve výpočtu nahrazeno nezáporným zbytkem při jeho dělení modulem. V modulární aritmetice s modulem 26 tak počítáme pouze s čísly 0, 1, ..., 25. Platí potom

$$17 + 19 = 10 \bmod 26,$$

protože $17 + 19 = 36$ a dělíme-li 36 modulem 26, dostaneme zbytek 10. Podobně

$$17 - 19 = 24 \bmod 26 \quad \text{a} \quad 7 \cdot 8 = 4 \bmod 26.$$

Dvě posloupnosti čísel stejné délky sčítáme (odčítáme) tak, že sečteme (odečteme) dvojice čísel na stejném místech posloupnosti.

Příklad 0.1 Sečtěte a odečtěte posloupnosti 15 11 23 06 11 a 17 04 14 19 23 modulo 26.

Řešení. Pro součet platí

$$\begin{array}{r}
 15 \quad 11 \quad 23 \quad 06 \quad 11 \\
 + \quad 17 \quad 04 \quad 14 \quad 19 \quad 23 \\
 \hline
 32 \quad 15 \quad 37 \quad 25 \quad 34 \\
 \text{mod } 26: \quad 6 \quad 15 \quad 11 \quad 25 \quad 9
 \end{array}$$

a pro rozdíl podobně

$$\begin{array}{r}
 15 \quad 11 \quad 23 \quad 06 \quad 11 \\
 - \quad 17 \quad 04 \quad 14 \quad 19 \quad 23 \\
 \hline
 -2 \quad 07 \quad 09 \quad -13 \quad -12 \\
 \text{mod } 26: \quad 24 \quad 07 \quad 09 \quad 13 \quad 14 \quad .
 \end{array}$$

□

Při počítání mod 10 používáme pouze čísla $0, 1, 2, \dots, 9$ a při počítání mod 2 počítáme pouze s čísly $0, 1$. Modulární aritmetika mod 2, neboli *binární aritmetika*, má tu vlastnost, že sčítání mod 2 se rovná odčítání mod 2:

$$\begin{array}{r}
 0 \quad 0 \quad 1 \quad 1 \\
 + \quad 0 \quad 1 \quad 0 \quad 1 \\
 \hline
 0 \quad 1 \quad 1 \quad 2 \\
 \text{mod } 2: \quad 0 \quad 1 \quad 1 \quad 0
 \end{array}$$

a

$$\begin{array}{r}
 0 \quad 0 \quad 1 \quad 1 \\
 - \quad 0 \quad 1 \quad 0 \quad 1 \\
 \hline
 0 \quad -1 \quad 1 \quad 0 \\
 \text{mod } 2: \quad 0 \quad 1 \quad 1 \quad 0 \quad .
 \end{array}$$

Modulární sčítání a odčítání písmen

Potřebujeme-li sčítat nebo odčítat posloupnosti písmen, používáme modulární aritmetiku s modulem 26. K tomu nahradíme každé písmeno dvouciferným číslem počínaje A=00 a konče Z=25 podle následující tabulky:

A	B	C	D	E	F	G	H	I	J	K	L	M
00	01	02	03	04	05	06	07	08	09	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Poté sečteme (odečteme) obě posloupnosti čísel mod 26 a nakonec výsledná čísla zpětně nahradíme písmeny podle stejné tabulky.

Příklad 0.2 *Sečtěte a odečtěte posloupnosti POSTEL a KOLENO.*

Řešení. Pro součet platí

$$\begin{array}{rcccccc}
 \text{POSTEL} & = & 15 & 14 & 18 & 19 & 04 & 11 \\
 + \text{KOLENO} & = & 10 & 14 & 11 & 04 & 13 & 14 \\
 \hline
 & & 25 & 02 & 03 & 23 & 17 & 25 & = & \text{ZCDXRZ}
 \end{array}$$

a rozdíl se rovná

$$\begin{array}{rcccccc}
 \text{POSTEL} & = & 15 & 14 & 18 & 19 & 04 & 11 \\
 - \text{KOLENO} & = & 10 & 14 & 11 & 04 & 13 & 14 \\
 \hline
 & & 05 & 00 & 07 & 15 & 17 & 23 & = & \text{FAHPRX.}
 \end{array}$$

□

Jednoduchá záměna

Caesarovy šifry a jejich řešení

Caesarova šifra zaměňuje každé písmeno písmenem, které je v abecedě o tři místa dále. Místo A píšeme D, písmeno B nahrazujeme písmenem E, atd. Julius Caesar si zvolil posunutí o tři místa, mohl ale zvolit posunutí o jakýkoliv počet míst mezi 1 a 25. Existuje tak 25 variant Caesarovy šifry. To je malý počet a pokud z nějakého důvodu víme, že použitá šifra pouze posouvá písmena abecedy o nějaký počet míst, můžeme při luštění postupovat tak, že vyzkoušíme všechny možnosti. Anglicky se tomuto postupu říká *exhaustive search*, česky budeme říkat, že šifru řešíme *hrubou silou*.

Z válečného tažení mohl Caesar poslat zprávu OXGB OBWB OBVB. Protože můžeme rozumně předpokládat, že asi použil nějakou variantu Caesarovy šifry, vyzkoušíme postupně všechna možná posunutí.

Posunutí	Zpráva
0	OXGB OBWB OBVB
1	PYHC PCXC PCWC
2	QZID QDYD QDXD
3	RAJE REZE REYE
4	SBKF SFAF SFZF
5	TCLG TGBG TGAG
6	UDMH UHCH UHBH
7	VENI VIDI VICI

Použitá varianta Caesarovy šifry proto velmi pravděpodobně posouvala každé písmeno abecedy o 19 míst dopředu, protože šifrový text dostaneme z otevřeného textu posunutím o 7 míst zpět a $19 = 26 - 7$. Pokud předpokládáme, že žádné další posunutí textu šifrové zprávy nedává smysluplný otevřený text, tak jsme zprávu správně rozluštili a nemusíme už zkoušet další posunutí. Tento předpoklad je rozumný v případě, že šifrová zpráva je dostatečně dlouhá. Krátká šifrová zpráva může mít více řešení. Například zpráva MSG vytvořená Caesarovou šifrou má nejméně dvě smysluplná řešení.

Posunutí	Zpráva
0	MSG
2	OUI
12	YES

Jednoduchá záměna

Při *jednoduché záměně* nahrazujeme normální abecedu nějakou její permutací. Každé písmeno normální abecedy při šifrování nahradíme, kdykoliv se objeví, písmenem které leží na stejném místě v permutované abecedě. Můžeme například použít následující permutaci.

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
Y	M	I	H	B	A	W	C	X	V	D	N	O	J	K	U	Q	P	R	T	F	E	L	G	Z	S

Pokud použijeme jednoduchou záměnu určenou právě uvedenou permutací abecedy a zašifrujeme zprávu

PRIJD VECER K ZELENEMU STROMU

dostaneme šifrový text

UPXVH EBIBP D SBNBJBOF RTPKOF

Pokus rozluštit jej stejně jako Caesarovu šifru k úspěchu nevede.

Jak by asi postupoval kryptoanalytik, kdyby se rozhodl, že text bude luštit jako jednoduchou záměnu? Využije základní slabinu jednoduché záměny, totiž že šifrový text má stejnou strukturu frekvence a rozmístění jednotlivých hlásek jakou má přirozený jazyk. Všimnul by si proto, že text se skládá z pěti slov, která mají délku pořadě 5, 5, 1, 8 a 6 písmen. Také by spočítal, že písmeno B se v šifrovém textu vyskytuje pětkrát na místech 7, 9, 13, 15 a 17, písmeno P třikrát na místech 2, 10 a 22, písmeno O dvakrát na místech 18 a 24, a písmeno F dvakrát na místech 19 a 25. Ostatní písmena se vyskytují každé pouze jednou. Celkem má text 25 písmen a z nich je 17 navzájem různých. Z toho vyplývá, že libovolný text v jakémkoliv jazyce používajícím mezinárodní abecedu, který má uvedené vlastnosti, je možným řešením tohoto šifrového textu. Například BZUCI LOMOZ K POHODOVE STRAVE nebo třeba MRAVY POZOR S KOLOTOCI NEBUCI. Ani jedna možnost sice nevypadá příliš pravděpodobně, nicméně jsou to také správná rozluštění textu UPXVH EBIBP D SBNBJBOF RTPKOF vytvořeného jednoduchou záměnou. To nás vede k přirozené otázce “Jak dlouhý musí být šifrový text, aby existovalo jediné řešení?”. V případě použití jednoduché záměny by mělo stačit asi tak 50 písmen, nemusí být ale jednoduché takový text rozluštit. Zkušenosti ukazují, že zhruba 200 písmen stačí k tomu, aby bylo také snadné šifrový text rozluštit.

Luštění zprávy UPXVH EBIBP D SBNBJBOF RTPKOF je výrazně usnadněné tím, že šifrový text obsahuje mezery mezi jednotlivými slovy. Tím okamžitě známe délky slov v otevřeném textu. Existují dva standardní způsoby, jak

tuto slabost jednoduché záměny odstranit. První způsob spočívá v ignorování mezer a interpunkčních znamének mezi jednotlivými slovy. V takovém případě otevřený text napíšeme jednoduše jako posloupnost písmen. Otevřené pozvání k návštěvě známé restaurace v Karlíně a jeho šifrovou verzi pak zapíšeme

PRIJDVE CERKZELENEMUSTROMU
UPXVHEBIBPDSBNBJBOFRTPKOF

Výsledkem je, že kryptoanalytik potom neví, z kolika slov jaké délky se původní otevřený text skládá. To samozřejmě zvyšuje počet možných řešení. Nevýhodou je, že také adresát zprávy při dešifrování musí do textu vložit mezery podle svého uvážení, což může vést k nejednoznačnosti, jak se snadno přesvědčíme na příkladu OKOLO TOC a O KOLOTOC. Úkol rozluštit šifrový text je tak těžší jak pro příjemce tak i pro kryptoanalytika.

Druhou používanou možností je nahradit v otevřeném textu každou mezeru nějakým málo užívaným písmenem, například písmenem X. V těch řídkých případech, kdy se v otevřeném textu objeví písmeno X, jej nahradíme vhodnou skupinou písmen, například bigramem KS. Naše zpráva a její šifrová verze potom vypadají následovně.

PRIJDXVECERKXKZELENEMUXSTROMU
UPXVHGEBIBPGDGSBNBJBOFGRTPKOF

V okamžiku, kdy kryptoanalytik přijde na to, že písmeno G znamená mezeru, tak nalezne délky jednotlivých slov. V případě delších zpráv je snadné symbol nahrazující mezeru poznat, jak si brzo ukážeme. Adresát zprávy nyní nebude mít problém správně rozložit přijatý text do jednotlivých slov, snažší to má ale také kryptoanalytik.

Jinou možností je přidat k abecedě nějaké další symboly, které budou nahrazovat mezeru a další interpunkční znaménka jako je tečka a čárka. Můžeme například použít &, \$, %. Číslo vypisujeme slovy, můžeme ale také k abecedě přidat další symboly. Tyto symboly navíc sice mohou šifrovou zprávu vytvořenou jednoduchou záměnou učinit na první pohled méně srozumitelnou, ve skutečnosti ale bezpečnost šifrové zprávy příliš nezvýší.

Všimněte si také, že permutace, kterou jsme používali, ponechává dvě písmena – Q a T – nezměněná. To není nijak na závadu. Ve skutečnosti lze spočítat, že náhodně zvolená permutace abecedy bude s přibližně dvoutřetinovou pravděpodobností vždy obsahovat nějaké písmeno, které se nezmění. Tato vlastnost není žádnou specialitou abecedy o 26 písmenech. Skoro stejně pravděpodobné je, že dva náhodně zamíchané balíčky s 52 kartami budou obsahovat nějakou kartu na stejném místě.

Jak vyřešit jednoduchou záměnu

Existuje $26! = 26 \cdot 25 \cdot \dots \cdot 3 \cdot 2 \cdot 1 > 4 \cdot 10^{26}$ permutací abecedy s 26 písmeny. Počítač, který by vyzkoušel za vteřinu jednu miliardu, tj. 10^9 permutací, by potřeboval několik set miliónů let, aby vyzkoušel všechny možnosti. Metoda vyzkoušet všechny možnosti, která tak dobře fungovala v případě Caesarových šifer, je v případě jednoduché záměny k ničemu. Je třeba postupovat jinak.

Praktický postup při řešení jednoduché záměny spočívá v následujících krocích.

1. Spočítáme frekvenci jednotlivých písmen v šifrovém textu.
2. Pokusíme se identifikovat písmeno, které případně nahrazuje mezeru mezi jednotlivými slovy. Pokud není šifrový text příliš krátký, tak se nám podaří zjistit, jestli takové písmeno existuje, a které to v tom případě je. Průměrná délka slov v přirozených jazycích je obvykle něco mezi 5 a 6 písmeny. Pokud tedy nějaké písmeno nahrazuje mezeru mezi slovy, musí tvořit něco mezi 16% a 20% textu. Příliš mnoho jiných písmen se v textu obvykle tak často neobjevuje. Dále, je-li správný náš předpoklad, že nějaké písmeno zastupuje mezeru, objeví se v šifrovém textu vždy po několika jiných písmenech, mezi jeho jednotlivými výskyty nejsou žádné dlouhé mezery a nemůže se také objevit dvakrát po sobě.
3. Pokud se nám podařilo identifikovat symbol nahrazující mezeru, napíšeme si šifrový text tak, že symbol nahrazující mezeru skutečně touto mezerou nahradíme. Dostaneme tak text tvořený jednotlivými 'slovy', která mají stejnou délku a strukturu jako slova v otevřeném textu. Pokud se tedy v nějakém otevřeném slově nějaké písmeno vyskytuje třikrát, je také třikrát v jeho šifrové podobě. Slovo SBNBJBOF tak může odpovídat slovům ZELENEMU, KOLOTOCI, POHODOVA, NEVESELY, nemůže ale nahrazovat slovo ZMRZLINA.
4. Pokusíme se identifikovat písmena, která v šifrovém textu nahrazují některá z nejčastěji používaných písmen v přirozeném jazyce, jako jsou například písmena E, A, T, P, atd. Několik těchto nejčastěji používaných písmen vždy tvoří až 40% obvyklého textu, ve většině jazyků a případů je písmeno E zdaleka nejčastější. Následující tabulka ukazuje frekvence jednotlivých písmen v některých jazycích. Tuto tabulku je třeba chápat pouze jako pomůcku, frekvence jednotlivých písmen v konkrétních textech se může hodně lišit od frekvencí uvedených v tabulce.

Písmeno	Angl.	Franc.	Něm.	Češ.	Slov.
A	7,96	7,68	5,52	8,99	9,49
B	1,60	0,80	1,56	1,86	1,90
C	2,84	3,32	2,94	3,04	3,45
D	4,01	3,60	4,91	4,14	4,09
E	12,86	17,76	19,18	10,13	9,16
F	2,62	1,06	1,96	0,33	0,31
G	1,99	1,10	3,60	0,48	0,40
H	5,39	0,64	5,02	2,06	2,35
I	7,77	7,23	8,21	6,92	6,81
J	0,16	0,19	0,16	2,10	2,12
K	0,41	0,00	1,33	3,44	3,80
L	3,51	5,89	3,48	4,20	4,56
M	2,43	2,72	1,69	2,99	2,97
N	7,51	7,61	10,20	6,64	6,34
O	6,62	5,34	2,14	8,39	9,34
P	1,81	3,24	0,54	3,54	2,87
Q	0,17	1,34	0,01	0,00	0,00
R	6,83	6,81	7,01	5,33	5,12
S	6,62	8,23	7,07	5,74	5,94
T	9,72	7,30	5,86	4,98	5,06
U	2,48	6,05	4,22	3,94	3,70
V	1,15	1,27	0,84	4,50	4,85
W	1,80	0,00	1,38	0,06	0,06
X	0,17	0,54	0,00	0,04	0,03
Y	1,52	0,21	0,00	2,72	2,57
Z	0,05	0,07	1,17	3,44	2,72

Nejčastěji používaná písmena se objevují pravidelně jako nejčastější písmena v různých textech. Zato málo frekventovaná písmena žádnou velkou cenu pro luštění nemají, v některém textu se mohou objevit častěji, v jiném nemusí být vůbec. Frekvence jednotlivých písmen také závisí na tom, o jaký text jde. Odborný text obsahující mnoho speciálních termínů může mít frekvence jednotlivých písmen velmi posunuté. Text o stavbě atomu často užívající termíny **PROTON**, **ELEKTRON**, **NEUTRON** bude mít patrně o dost vyšší frekvenci **O**, než je obvyklé. Frekvence uvedené v tabulce vycházejí z textů v několika evropských jazycích obsahujících více než deset tisíc písmen a jsou uvedené v procentech.

Vzhledem k tomu, že nejdůležitější je vyhledat v šifrovaném textu pís-

mena odpovídající nejčastěji používaným písmenům v otevřeném textu, uvedeme také tabulku šesti nejčastěji používaných písmen v jednotlivých jazycích.

Angl.	Franc.	Něm.	Čeština	Slov.
E: 12,86	E: 17,76	E: 19,18	E: 10,13	A: 9,49
T: 9,72	S: 8,23	N: 10,20	A: 8,99	O: 9,34
A: 7,96	A: 7,68	I: 8,21	O: 8,39	E: 9,16
I: 7,77	N: 7,61	S: 7,07	I: 6,92	I: 6,81
N: 7,51	T: 7,30	R: 7,01	N: 6,64	N: 6,34
R: 6,83	I: 7,23	T: 5,86	S: 5,74	S: 5,94
Σ : 52,65	Σ : 55,81	Σ : 57,53	Σ : 46,81	Σ : 47,08

Také je dobré vědět, která písmena se nejčastěji vyskytují na začátku a na konci jednotlivých slov. Následující tabulka ukazuje nejčastější písmena na začátku a na konci slov v češtině. Byla vytvořena na základě 18 938 slov.

Začátek	Konec
P: 12,50	E: 16,67
S: 9,72	I: 13,96
V: 9,19	A: 10,94
Z: 8,95	O: 8,93
N: 7,64	U: 7,94
O: 5,56	Y: 7,03
Σ : 53,56	Σ : 65,47
souhl.: 84,51	souhl.: 34,53
samohl.: 15,49	samohl.: 65,47

5. Pokud jsme již tímto způsobem identifikovali části jednotlivých slov, hledáme krátká slova, ve kterých už nějaká písmena známe. V angličtině jsou například nejčastějšími písmeny E a T. Pokud najdeme v otevřeném textu slovo T.E, je velmi pravděpodobné, že jde o slovo THE. Podobně najdeme-li v českém textu slovo A.E, jde s velkou pravděpodobností o slovo ALE. Využijeme také informace o frekvencích bigramů v jednotlivých jazycích. Následující tabulka uvádí deset nejčastějších bigramů ve stejných jazycích (s výjimkou slovenštiny). Je třeba opět upozornit, že tyto tabulky jsou vytvářené na základě konkrétních textů

a frekvence jednotlivých bigramů v různých textech může být různá. Tabulka je tak další pomůckou při řešení jednoduché záměny.

Angl.	Franc.	Něm.	Čeština
TH: 3,30	ES: 3,05	EN: 4,43	PR: 1,98
HE: 2,70	EL: 2,46	ER: 3,75	NI: 1,94
IN: 2,02	EM: 2,42	CH: 2,80	ST: 1,81
ER: 1,91	DE: 2,15	EI: 2,42	NA: 1,68
RE: 1,69	RE: 2,09	DE: 2,33	NE: 1,61
AN: 1,67	NT: 1,97	ND: 2,08	EN: 1,55
ES: 1,49	ON: 1,64	IN: 1,97	RA: 1,35
EN: 1,46	ER: 1,63	GE: 1,96	OV: 1,32
ON: 1,34	TE: 1,63	IE: 1,88	TE: 1,30
AT: 1,27	SE: 1,55	TE: 1,76	AN: 1,25

V češtině a slovenštině se prakticky nevyskytují zdvojená stejná písmena, zatímco v angličtině, němčině a francouzštině jsou častá. Následující tabulka ukazuje devět nejčastějších zdvojenin v těchto jazycích.

Angl.	Franc.	Něm.
TT: 0,56	SS: 0,73	SS: 0,82
LL: 0,53	EE: 0,66	LL: 0,36
EE: 0,51	LL: 0,66	EE: 0,35
SS: 0,48	TT: 0,29	NN: 0,34
RR: 0,24	NN: 0,24	TT: 0,28
FF: 0,21	MM: 0,20	RR: 0,21
OO: 0,13	RR: 0,17	FF: 0,17
PP: 0,09	PP: 0,16	MM: 0,13
CC: 0,07	FF: 0,10	DD: 0,10

6. Nakonec dokončíme řešení s využitím gramatických pravidel a informací vyplývajících z kontextu.

Vybaveni těmito informacemi se nyní můžeme pokusit o řešení nějakého šifrovaného textu vytvořeného pomocí jednoduché záměny.

Příklad 1.1 Víme, že následující šifrový text byl vytvořen jednoduchou záměnou z anglického textu, a víme dále, že mezery v původním textu byly před zašifrováním nahrazené písmenem Z. Najděte otevřený text.

MJZYB LGESE CNCMQ YGXYS PYZDZ PMYGI IRLLC
 PAYCK YKGWZ MCWZK YFCRM ZYVCX XZLZP MYXLG
 WYMJS MYGPZ YWCAJ MYCWS ACPZY XGLYZ HSWBN
 ZYXZT YTGRN VYMJC POYMJ SMYCX YMJZL ZYSLZ
 YMTZP MQYMJ LZZYB ZGBNZ YCPYS YLGGW YMJZP
 YMJZL ZYCKY SPYZD ZPKYI JSPIZ YMJSM YMJZL
 ZYSLZ YMTGY GXYMJ ZWYTC MJYMJ ZYKSW ZYECL
 MJVSQ YERMY MJCKY CKYKG

Řešení.

1. V textu je 53 skupin po pěti písmenech, celkem 265 písmen. Spočítáme, kolikrát se které písmeno v šifrovaném textu vyskytuje.

A:	3	E:	4	I:	4	M:	27	Q:	3	U:	0	Y:	49
B:	4	F:	1	J:	17	N:	4	R:	4	V:	3	Z:	33
C:	18	G:	14	K:	9	O:	1	S:	14	W:	9		
D:	2	H:	1	L:	14	P:	13	T:	6	X:	8		

2. Nejčastěji se vyskytuje písmeno Y, celkem 49x, což je zhruba 18,5% celého textu. Písmeno Y je tak dobrým kandidátem pro mezeru v šifrovaném textu. Dalšími dvěma nejčastějšími písmeny jsou Z a M. To jsou vhodné kandidáti na nejčastější písmena v anglických textech E, T nebo (méně pravděpodobně) T, E.
3. Nyní nahradíme v šifrovaném textu písmeno Y mezerou. Ignorujeme mezery mezi jednotlivými pěticemi v šifrovaném textu, které nemají žádný význam. Dostaneme tak text, který odhaluje délky slov. Těch je celkem 50. Jednotlivá slova si očíslováme, abychom na ně mohli při dalším řešení odkazovat. V textu je poměrně dost krátkých slov, průměrná délka slova je o něco více než 5 písmen.

1	2	3	4	5	6	7	
MJZ	BLGESECNCMQ	GX	SP	ZDZPM	GIIRLLCPA	CK	
8	9	10	11	12	13	14	
KGWZMCWZK	FRCMZ	VCXXZLZPM	XLGW	MJSM	GPZ	WCAJMJ	
15	16	17	18	19	20	21	22
CWSACPZ	XGL	ZHSWBNZ	XZT	TGRNV	MJCPO	MJSM	CX
23	24	25	26	27	28	29	30
MJZLZ	SLZ	MTZPMQ	MJLZZ	BZGBNZ	CP	S	LGGW

31	32	33	34	35	36	37	38	
MJZP	MJZLZ	CK	SP	ZDZPK	IJSPIZ	MJSM	MJZLZ	
39	40	41	42	43	44	45	46	47
SLZ	MTG	GX	MJZW	TCMJ	MJZ	KSWZ	ECLMJVSQ	ERM
48	49	50						
MJCK	CK	KG						

Rozložení délky slov odpovídá zhruba délkám slov v přirozeném jazyce, což dále podporuje naši hypotézu, že písmeno Y v šifrovém textu odpovídá mezeře v otevřeném textu.

4. Nyní se podíváme na krátká slova v šifrovém textu.

- Slovo **S** s číslem **29** má délku 1. Odhadneme proto, že šifrové **S** je pravděpodobně otevřené **A** nebo **I**.
- Deset slov má délku 2. Z toho se **CK** vyskytuje třikrát na místech **7**, **33** a **49**, a dvě slova se objevují dvakrát – slovo **GX** na místech **3** a **41** a slovo **SP** na místech **4** a **34**.
- Jedenáct slov má délku 3, dvě z nich se objevují dvakrát – slovo **MJZ** na místech **1** a **44** a slovo **SLZ** na místech **24** a **39**.

5. Protože už máme podezření, že písmena **M, Z** v šifrovém textu jsou patrně otevřená písmena **T, E** nebo naopak **E, T**, tak vidíme, že trigram **MJZ** je buď **T?E** nebo **E?T**, a protože se objevuje dvakrát, tak je velmi pravděpodobné, že je to **THE**. V šifrovém textu tak písmena **M, J** a **Z** odpovídají otevřeným písmenům **T, H** a **E**. Existuje ještě několik dalších slov, ve kterých se vyskytují písmena **M, Z** a **J**. Jsou to

- číslo **23** – slovo **MJZLZ**, což je tedy **THE?E**, čili **L** znamená buď **R** nebo **S**,
- číslo **26** – slovo **MJLZZ**, což je **TH?EE**, čili **L** znamená **R**,
- číslo **42** – slovo **MJZW**, což je **THE?**, a tak **W** je znamená buď **M** nebo **N**,
- číslo **37** – slovo **MJSM**, neboli **TH?T**, a protože už víme, že **S** znamená buď **A** nebo **I**, odpovídá šifrové slovo **MJSM** buď otevřenému **THAT** nebo **THIT**.

Z těchto úvah tak vyplývá, že **S** je otevřené **A**, **L** je otevřené **R** a **W** nahrazuje buď otevřené **M** nebo otevřené **N**.

Slovo **26** se ukázalo být slovem **THREE**. Podíváme se proto na slovo **25**, jestli není náhodou také nějakým číslem. Zatím víme, že se rovná otevřenému **T?E?T?**, což nápadně připomíná slovo **TWENTY**. Pokud tomu tak je, dostáváme, že písmena **T**, **P** a **Q** v šifrovém textu odpovídají písmenům **W**, **N** a **Y** v otevřeném textu. Tím by také byla vyřešena nejistota týkající se šifrového **W**, které by tak muselo odpovídat otevřenému **M**.

6. Zjistili jsme tak, že devíti šifrovým písmenům **J**, **L**, **M**, **P**, **Q**, **S**, **W**, **Y** a **Z** odpovídají v otevřeném textu **H**, **R**, **T**, **N**, **Y**, **A**, **M**, mezera a **E**. Těchto devět písmen tvoří více než 60% textu. Napíšeme si znovu šifrový text a pod něj odpovídající písmena otevřeného textu, pokud je už známe. Dále napíšeme tečku . tam, kde ještě otevřené ekvivalenty šifrových písmen neznáme.

To nás přivede k několika dalším písmenům. Tak například šifrové slovo **LGGW** (číslo **30**) je **R. .M**, přičemž uprostřed je dvojice stejných písmen. To dává jedinou možnost pro otevřený text, slovo **ROOM**. Šifrové **G** je tedy otevřené **O**. Slova číslo **48** a **49** jsou **MJCK** a **CK** a ta jsme již částečně rozluštili jako **TH. S** a **. S**. šifrové **C** proto odpovídá otevřenému **I**. Poslední tři šifrová slova **MJCK CK KG** tak rozluštíme jako **THIS IS ?O**, neboť už víme, že šifrovému **G** odpovídá otevřenému **O**. Proto je šifrové slovo **KG** otevřené **SO**, neboli šifrové **K** odpovídá otevřenému písmenu **S**.

Dosadíme tak dále za šifrová písmena **C**, **K** a **G** pořadě otevřená písmena **I**, **O** a **S**. Částečně rozluštěný text tak vypadá následovně.

1	2	3	4	5	6	7	
MJZ	BLGESECNCMQ	GX	SP	ZDZPM	GIIRLLCPA	CK	
THE	.RO.A.I.ITY	O.	AN	E.ENT	O...RRIN.	IS	
8	9	10	11	12	13	14	
KGWZMCWZK	FRCMZ	VCXXZLZPM	XLGW	MJSM	GPZ	WCAJM	
SOMETIMES	..ITE	.I..ERENT	.ROM	THAT	ONE	MI.HT	
15	16	17	18	19	20	21	22
CWSACPZ	XGL	ZHSWBNZ	XZT	TGRNV	MJCPO	MJSM	CX
IMA.INE	.OR	E.AM..E	.E.	WO...	THIN.	THAT	I.
23	24	25	26	27	28	29	30
MJZLZ	SLZ	MTZPMQ	MJLZZ	BZGBNZ	CP	S	LGGW
THERE	ARE	TWENTY	THREE	.EO..E	IN	A	ROOM
31	32	33	34	35	36	37	38
MJZP	MJZLZ	CK	SP	ZDZPK	IJSPIZ	MJSM	MJZLZ
THEN	THERE	IS	AN	E.ENS	..AN.E	THAT	THERE

39	40	41	42	43	44	45	46	47
SLZ	MTG	GX	MJZW	TCMJ	MJZ	KSWZ	ECLMJVSQ	ERM
ARE	TWO	O.	THEM	.ITH	THE	SAME	.IRTH.AY	. .T
48	49	50						
MJCK	CK	KG						
THIS	IS	SO						

Nyní už snadno doplníme zbývající šifrová písmena jejich otevřenými ekvivalenty. Například ze slov 14 a 15 vyplývá, že šifrové A odpovídá otevřenému G. Slovo 20 pak znamená, že šifrové O odpovídá jednomu z otevřených písmen G nebo K. Protože G je už obsazené, a také z kontextu, dostáváme že šifrové O je otevřené K. Ze slova 41 vyplývá, že šifrové X může odpovídat jednomu z otevřených písmen F, N nebo R. Ani N ani R to už být nemůže, proto musí odpovídat otevřenému F, atd. Dostaneme tak *dešifrovací* abecedu.

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
G	P	I	V	B	Q	O	X	C	H	S	R	T	L	K	N	Y	U	A	W	.	D	M	F	_	E

Podtržítka _ označuje mezeru mezi slovy. *Šifrovací* abeceda, která byla použita při šifrování otevřeného textu, je samozřejmě inverzní permutace k dešifrovací abecedě:

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
S	E	I	V	Z	X	A	J	C	.	O	N	W	P	G	B	F	L	K	M	R	D	T	H	Q	Y

Otevřený text je tak

THE PROBABILITY OF AN EVENT OCCURING IS
SOMETIMES QUITE DIFFERENT FROM WHAT ONE MIGHT
IMAGINE FOR EXAMPLE FEW WOULD THINK THAT IF
THERE ARE TWENTY THREE PEOPLE IN A ROOM
THEN THERE IS AN EVENS CHANCE THAT THERE
ARE TWO OF THEM WITH THE SAME BIRTHDAY BUT
THIS IS SO

Písmeno U se nevyskytuje v šifrovém textu, zatímco písmena J a Z se nevyskytují v otevřeném textu. Písmeno Z označuje mezeru v otevřeném textu a bylo zašifrováno jako Y. Na otevřené písmeno J tak zbývá jediné písmeno, které v šifrovém textu neexistuje, písmeno U. □

Autorem následujícího příkladu je Mgr. Pavel Vondruška. Ještě než se pustíme do jeho řešení, uvedeme si několik dalších zvláštností českého jazyka. Samohlásky a souhlásky se v českých slovech víceméně střídají, písmena R a L v některých případech vystupují v roli samohlásek. Nejfrekventovanější samohlásky jsou pořadě E, A, O, I, samohlásky U a Y jsou mnohem méně

frekventované. V českých slovech se prakticky nevyskytují dvojice po sobě jdoucích samohlásek. Jedinou výjimkou je bigram OU.

Pokud při šifrování vynecháme mezery, může se stát, že jedno slovo končí samohláskou a druhé samohláskou začíná. Dvě třetiny českých slov sice na samohlásku končí, ale pouze méně než 15% českých slov samohláskou začíná. Dvě samohlásky vedle sebe se tak vyskytují i v takovém případě jen zřídka. Můžeme také odhadnout, jak často. Průměrná délka českých slov je 5,37 písmen. Počet slov v nějakém textu je tak méně než jedna pětina délky textu. Přesně tolik bigramů tak odpovídá poslednímu písmenu jednoho slova a počátečnímu písmenu následujícího slova. Z těchto bigramů je $0,15 \cdot 0,66 \doteq 0,1$ takových, že jedno slovo končí samohláskou a následující samohláskou začíná. Proto zhruba $(1 : 5,37) \cdot 0,15 \cdot 0,66 \doteq 0,0184$ bigramů je tvořeno dvojicí samohlásek, z nichž jedna je na konci jednoho slova a druhá na počátku následujícího. Přidáme součet frekvencí samohláskových bigramů v jednotlivých slovech, který je zhruba 1,5%. Dohromady tedy samohláskové bigramy tvoří přibližně 2,5% všech bigramů.

Tím se dostáváme ke zvláštnostem některých frekventovaných bigramů.

1. bigram OU:

- jde o zdaleka nejfrekventovanější samohláskový bigram, zhruba 0,77%, zatímco frekvence samostatného O je 8,39%, více než jedenáctkrát větší. Frekvence samotného U je 3,94, přibližně pětkrát větší než frekvence OU a poloviční oproti frekvenci samotného O,
- zatímco bigram OU patří mezi pět vůbec nejčastějších bigramů v češtině, obrácený bigram UO se prakticky nevyskytuje.

2. bigram ST:

- písmena S a T mají přibližně stejnou frekvenci,
- existuje i bigram TS, ten se ale vyskytuje s frekvencí víc než třicetkrát menší, než je frekvence ST,
- je součástí velkého počtu hodně frekventovaných souhláskových trigramů STR, STN, STL, STV, atd.,
- vyskytuje se uprostřed i na konci slov.

3. bigram PR:

- písmeno P má přibližně poloviční frekvenci než písmeno R,
- obrácený bigram RP se prakticky nevyskytuje (jednou z výjimek je CHRPA),

- zpravidla stojí na počátku slov,
- často lze doplnit na hláskové trigramy SPR, ZPR
- jen zřídka lze doplnit na hláskový trigram PRV a pokud ano, tak téměř výhradně na začátku slov.

4. bigram CH:

- písmeno H má frekvenci přibližně 2%, písmeno C má frekvenci přibližně 3%, a celý bigram CH má frekvenci přibližně 1%,
- opačný bigram HC se prakticky nevyskytuje,
- bývá zpravidla na konci slov spolu se samohláskami Y, A, E, I,
- obvykle platí, že předchází-li před CH souhláska, následuje po něm samohláska, a naopak (příklady: OBCHOD, NECHTĚL, atd.).

Nejčastějšími trigramy v českých textech jsou PRO, OVA, ENI, PRI, OST, PRA, ANI, STA, atd. Všechny s frekvencemi od 0,8% do 0,5%. Zdaleka nejfrekventovanějšími souhláskovými trigramy jsou STR s frekvencí 0,24% a STN s frekvencí 0,16%.

Příklad 1.2 *Následující šifrový text je vytvořený jednoduchou záměnou z českého textu napsaného v mezinárodní abecedě (bez diakritických znamének) a bez mezer. Najděte příslušný otevřený text.*

```
UFTAL OTCSF CILDO TGLUL JHSFN PZIHf NGBZU FTALP
ZRZOB NCHSF NQBZA ZFZGX ZWOZG OLPZX AHBHU FTALP
ZXIHJ OTWZJ HFAZD NDTOS BZLFN WCHPR ZPHCI TUXHI
ZCITD ZSAWT BCHSF NDNFT ALPZG ZGZPZ WIZD NQAHS
WZOTP TCOZJ RZHWT UBTPZ HJOZW TUBHB LHJUB ALOTP
ZWLUB TOLXL JZOZI LADLP TCPNG SGDNU ZOLOL ULQIT
DHUBX IHJOT WYDHJ HSRZG OLPQH SGZXI HJOTW ZRZXA
ZUBLA ILOZD CHJOL QZUFZ ASXAT AHGQI LJONW CXAHW
ZUZWC PHCHS DGOTQ LBTOZ FZGXZ WOZIL BQNRL QHOLX
ARZJO ZSATO BLQUZ PHCHS UBLBT BGDRZ JIZCH SFNJA
SCHBO ZRZJH DLBNP TDNRP SBZXI HJOTW ZSQIL JLPZJ
HHBZD AZONW CJNWC LRTWT WCHFL ISOZF HBDSG LDAZO
NWCOZ XAHXS UBONW CHFLI ZWCUZ XIHJO TWZBG DGLXL
ATGDI LUBZO ZDCHJ OZRZS IHGZO TDXHI NZBNI ZOHDN
WCULW WTWCO ZRDCH JOZRU TPTHF LINGS UBLDL RTBAL
JTWoz XIZBZ OZQHU TWQNO ZRXHG JZRTJ ASCNJ ZOxHU
FZASP LRtFN BCHSF NGXAL WHDLO NEEEE
```

Řešení. Celý text má 670 písmen. Spočítáme absolutní a relativní frekvence nejčastějších pěti písmen:

Písmeno	Počet	%
Z	82	12,24
H	52	7,76
L	48	7,16
O	46	6,87
T	42	6,27

Dalším nejfrekventovanějším písmenem je B s 32 výskyty, což je 4,78%. To je o dost menší než výskyty více frekventovaných písmen, proto zkusíme, jestli písmena Z, H, L, O a T v šifrovém textu nenahrazují nejčastěji používaná písmena v otevřených českých textech E, A, O, I a N.

Pokusíme se odhalit, která z písmen Z, H, L, O a T v šifrovém textu mohou zastupovat souhlásky. K tomu spočítáme výskyty všech 25 možných bigramů složených z těchto písmen. Vidíme, že všechny nejfrekventovanější bigramy obsahují písmeno O, zatímco ostatní bigramy se vyskytují pouze zřídka. Odtud usoudíme, že písmeno O nahrazuje souhlásku, zatímco ostatní nahrazují samohlásky. Zdaleka nejčastějším písmenem v otevřených českých textech je E, odhadneme proto, že je v šifrovém textu nahrazeno nejfrekventovanějším písmenem Z. Zbývající tři šifrová písmena H, L, T tak pravděpodobně zastupují otevřené samohlásky A, O, I.

V první fázi nám jde především o odhalení samohlásek. Pro čtyři z nich už máme pravděpodobné kandidáty Z, H, L, T. Napišeme si je proto pod příslušná místa do několika prvních řádků šifrového textu. S výjimkou dvojice Z a E nevíme, která otevřená samohláska odpovídá kterému z písmen H, L, T v šifrovém textu. V této chvíli na tom ale tolik nezáleží. Zkusíme je proto nahradit pořadě samohláskami A, O, I, jak to odpovídá frekvencím těchto samohlásek v českých textech. Dostaneme tak

UFTAL	OTCSF	CILDO	TGLUL	JHSFN	PZIHf	NGBZU	FTALP
..I.O	NI...	..O.N	I.O.O	A....	.E.A.	...E.	..I.O
ZRZOB	NCHSF	NQBZA	ZFZGX	ZWOZG	OLPZX	AHBHU	FTALP
E.EN.	..A..	E....	E.E..	..NE.	NO.E.	.A.A.	.I.O.
ZXIHJ	OTWZJ	HFAZD	NDTOS	BZLFN	WCHPR	ZPHCI	TUXHI
E..O.	NI.E.	A..E.	..IN.	.EO..	..A..	E.A..	I..A.
ZCITD	ZSAWT	BCHSF	NDNFT	ALPZG	ZGZPZ	WZIZD	NQAHS
E..I.	E...I	..A..	I	.O.E.	E.E.E	.E.E.	...A.

WZOTP	TCOZJ	RZHWT	UBTPZ	HJOZW	TUBHB	LHJUB	ALOTP
..NIP	I.NE.	.EA.I	..I.E	A.NE.	I..A.	OA...	.ONI.
ZWLUB	TOLXL	JZOZI	LADLP	TCPNG	SGDNU	ZOLOL	ULQIT
E.O..	INO.O	.ENE.	O..O.	I....		ENONO	.O..I

Zběžný pohled ukazuje, že rozmístění samohlásek E, A, O, I odpovídá rozmístění samohlásek v českém textu. Ani poloha písmene N není nikde ve zřejmém rozporu s pravidly českého pravopisu.

Otevřený text stále ještě obsahuje mnoho, celkem 29, polygramů délky aspoň 4 s celkovým počtem 150 písmen. Spočítáme, která písmena se v těchto polygramech vyskytují nejčastěji:

Písmeno	Počet
N	24
S	15
C	13
B	13

Ostatní písmena se v těchto vybraných polygramech vyskytují nejvýše 10x. Protože pátráme po dvou písmenech v šifrovém textu, která zastupují zbývajících dvě otevřené samohlásky U a Y, podíváme se, ve kterých polygramech tvořených aspoň pěti písmeny schází každá dvojice (bez ohledu na pořadí) vytvořená z písmen N, S, C, B.

Dvojice	Schází v
N, S	–
N, C	–
N, B	CSFCI
S, C	ONEEEE
S, B	JONWCXA, ONWCJNWC, DNWCU, ONEEEE
C, B	SFNDNF, PNGSGDNU, SFNJAS, INGSU, SFNGXA, ONEEEE

Závěrečný polygram šifrového textu ONEEEE je zvláštní tím, že obsahuje čtyři sousední stejná písmena. To naznačuje, že čtveřice EEEE bylo k textu přidána tak, aby výsledný počet písmen byl násobkem pěti. Tento polygram tedy nebudeme uvažovat. Dvojice C, B schází v několika různých dlouhých polygramech, jednom délky 8 a třech délky 6. Tak dlouhé polygramy bez samohlásek se v češtině vyskytují jen velmi zřídka. Nejpravděpodobnější tak je, že v šifrovém textu písmena N a S zastupují zbývajících dvě otevřené

samohlásky U a Y. O něco méně pravděpodobnější je, že posledním dvěma samohláskám odpovídá jedna z dvojic N, C nebo S, C.

Vyzkoušíme tedy, že samohláskám v otevřeném textu odpovídají v šifrovaném textu písmena Z, H, L, T, N a S. Nyní spočítáme výskyt všech možných bigramů z těchto šesti písmen. Jeden z nich, bigram HS, se v šifrovaném textu vyskytuje 10x, zatímco všechny ostatní bigramy z těchto šesti písmen se vyskytují nejvýše dvakrát. To napovídá, že šifrový bigram HS odpovídá otevřenému samohláskovému bigramu OU. Tím dostáváme následující tabulku pro dešifrování samohlásek:

Z	H	L	T	N	S
E	O	A	I	Y	U

Význam písmen H a L v šifrovaném textu je tedy opačný, než jak jsme dosud předpokládali. Zkusíme tedy dosadit podle této tabulky do několika počátečních řádků šifrovaného textu. Dostaneme tak

UFTAL	OTCSF	CILDO	TGLUL	JHSFN	PZIHf	NGBZU	FTALP
..I.A	NI.U.	..A.N	I.A.A	O.U.Y	.E.O.	Y..E.	.I.A.
ZRZOB	NCHSF	NQBZA	ZFZGX	ZWOZG	OLPZX	AHBHU	FTALP
E.EN.	Y.OU.	Y..E.	E.E..	..NE.	NA.E.	.O.O.	.I.A.
ZXIHJ	OTWZJ	HFAZD	NDTOS	BZLFN	WCHPR	ZPHCI	TUXHI
E..A.	NI.E.	O..E.	Y.INU	.EA.Y	..O..	E.O..	I..O.
ZCITD	ZSAWT	BCHSF	NDNFT	ALPZG	ZGZPZ	WZIZD	NQAHS
E..I.	EU..I	..OU.	Y.Y.I	.A.E.	E.E.E	.E.E.	Y..OU
WZOTP	TCOZJ	RZHWT	UBTPZ	HJOZW	TUBHB	LHJUB	ALOTP
..NIP	I.NE.	.EO.I	..I.E	O.NE.	I..O.	AO...	.ANI.
ZWLUB	TOLXL	JZOZI	LADLP	TCPNG	SGDNU	ZOLOL	ULQIT
E.A..	INA.A	.ENE.	A..A.	I..Y.	U..Y.	ENANA	.A..I

Rozložení samohlásek v textu vypadá přijatelně pro český text. Budeme tedy předpokládat, že samohlásky jsme už správně odhalili a budeme se věnovat souhláskám. Najdeme nejčastější souhláskové bigramy.

Bigram	Počet	První	Druhá
WC	11	31	28
UB	10	25	32
JO	9	27	46
XA	7	22	28
UF	5	25	22

Všechny ostatní souhláskové bigramy se v šifrovém textu vyskytují méně než pětikrát. Mezi nejčastější souhláskové bigramy v českých textech patří bigram CH, přičemž souhláska C je přibližně třikrát častější a souhláska H je zhruba dvakrát častější než celý bigram CH. Těmto poměrům nejvíce odpovídá bigram WC v šifrovém textu. Zkusíme tedy dosadit místo šifrového W otevřené C a místo šifrového C otevřené H. Mezi nejčastějšími souhláskovými bigramy v českých textech se vyskytuje několik, které mají na prvním místě písmeno S, zatímco jiné souhlásky se na prvním místě nejfrekventovanějších bigramů objevují pouze jednou. V našem přehledu pěti nejčastějších bigramů se objevuje jedno šifrové písmeno dvakrát, a to U. Zkusíme tedy dále předpokládat, že šifrové U odpovídá otevřenému S. V bigramu JO známe druhé písmeno, neboť O odpovídá nejčastější souhlásce, kterou je N. Nejčastějším bigramem v českých textech, který má na druhém místě písmeno N, je bigram DN. Asi o polovinu méně častější, přesto dosti frekventované, jsou také bigramy TN a ZN. Zkusíme proto ještě dosadit za šifrové J otevřené D. Zopakujme si všechna šifrová písmena, která jsme dosud zkusili nahradit odpovídajícími písmeny otevřeného textu.

Z H L T N S O W C U J
E O A I Y U N C H S D

Počet výskytů písmen v prvním řádku v šifrovém textu je celkem 436, což jsou téměř dvě třetiny celého textu. Pokud jsme se nezmýlili, pak by měla odhalená písmena stačit k jednoduchému dolůštění celého textu. Tak to vyzkoušíme.

UFTAL	OTCSF	CILDO	TGLUL	JHSFN	PZIHf	NGBZU	FTALP
S.I.A	NIHU.	H.A.N	I.ASA	DOU.Y	.E.O.	Y..ES	.I.A.
ZRZOB	NCHSF	NQBZA	ZFZGX	ZWOZG	OLPZX	AHBHU	FTALP
E.EN.	YHOU.	Y..E.	E.E..	.CNE.	NA.E.	.O.OS	.I.A.
ZXIHJ	OTWZJ	HFAZD	NDTOS	BZLFN	WCHPR	ZPHCI	TUXHI
E..A.	NI.ED	O..E.	Y.INU	.EA.Y	CHO..	E.OH.	IS.O.
ZCITD	ZSAWT	BCHSF	NDNFT	ALPZG	ZGZPZ	WZIZD	NQAHS
EH.I.	EU.CI	.HOU.	Y.Y.I	.A.E.	E.E.E	CE.E.	Y..OU
WZOTP	TCOZJ	RZHWT	UBTPZ	HJOZW	TUBHB	LHJUB	ALOTP
CENI.	IHNED	.EOCI	S.I.E	ODNEC	IS.O.	AODS.	.ANI.
ZWLUB	TOLXL	JZOZI	LADLP	TCPNG	SGDNU	ZOLOL	ULQIT
ECAS.	INA.A	DENE.	A..A.	IH.Y.	U..YS	ENANA	SA..I

Nyní se podíváme na následující část textu tvořenou předposledním řádkem:

WZOTP	TCOZJ	RZHWT	UBTPZ	HJOZW	TUBHB	LHJUB	ALOTP
CENI.	IHNED	.EOCI	S.I.E	ODNEC	IS.O.	AODS.	.ANI.

V otevřeném druhém řádku jsou hned dva výskyty samohláskového bigramu EO, který naznačuje, že jde o poslední písmeno jednoho slova a první písmeno následujícího slova. Těsně před prvním výskytem je otevřené slovo IHNED, což znamená, že pro šifrový bigram RZ, který pravděpodobně odpovídá slovu otevřeného textu o dvou písmenech, máme z kontextu, ve kterém se nachází, pouze dvě možnosti: JE a NE, přičemž ta druhá je vyloučena, neboť otevřené N je šifrováno pomocí O. Proto písmeno R odpovídá otevřenému J. Z následujících čtyř neznámých otevřených písmen jsou tři šifrována stejným písmenem B. Vyzkoušením všech možných souhlásek dostaneme, že jedinou vhodnou možností je otevřené písmeno T. Potom P zjevně šifruje otevřené písmeno M. Doplňme je do dosud nalezené tabulky pro luštění šifrovaného textu.

Z	H	L	T	N	S	O	W	C	U	J	P	B	R
E	O	A	I	Y	U	N	C	H	S	D	M	T	J

Po doplnění těchto tří písmen je předposlední řádek následující:

WZOTP	TCOZJ	RZHWT	UBTPZ	HJOZW	TUBHB	LHJUB	ALOTP
CENIM	IHNED	JEOCI	STIME	ODNEC	ISTOT	AODST	.ANIM

Proto šifrové A odpovídá otevřenému R. Takto postupně doplníme celou tabulku pro luštění, dešifrování textu. První řádek srovnáme podle abecedy, aby se v něm lépe hledalo.

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
L	F	W	J	Z	K	V	C	T	R	Q	I	P	O	H	X	Y	A	U	B	S	D	M	E	N	G

Na vhodná místa podle smyslu dosadíme mezery a celý otevřený text pak vypadá následovně.

SBIRANI HUB HLAVNI ZASADOU BY MELO BYT ZE SBIRAME JEN TY
 HOUBY KTERE BEZPECNE ZNAME PROTO SBIRAME PLODNICE DOBRE
 VYVINUTE ABYCHOM JE MOHLI SPOLEHLIVE URCIT HOUBY VYBIRAME
 ZE ZEME CELE VYKROUCENIM IHNED JE OCISTIME OD NECISTOT A
 ODSTRANIME CASTI NAPADENE LARVAMI HMYZU ZVYSENA SAKLIVOST
 PLODNICE VODOU JE ZNAMKOU ZE PLODNICE JE PRESTARLA NEVHODNA
 KE SBERU PRI ROZKLADNYCH PROCESECH MOHOU VZNIKAT I
 NEBEZPECNE LATKY JAKO NAPR JED NEURIN TAK SE MOHOU STAT I
 TZV JEDLE HOUBY DRUHOTNE JEDOVATYMI VYJMUTE PLODNICE

UKLADAME DO OTEVRENYCH DYCHAJICICH OBALU NEBOT V UZAVRENYCH
NEPROPUSTNYCH OBALECH SE PLODNICE TZV ZAPARI ZVLASTE
NEVHODNE JE ULOZENI V POLYETYLENOVYCH SACCICH
NEJVHODNEJSIMI OBALY ZUSTAVAJI TRADICNE PLETENE KOSICKY
NEJPOZDEJI DRUHY DEN PO SBERU MAJI BYT HOUBY ZPRACOVANY
XXXX

K šifrování tohoto otevřeného textu byla použita následující tabulka.

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
R	T	H	V	X	B	Z	O	L	D	F	A	W	Y	N	M	K	J	U	I	S	G	C	P	Q	E

□

Luštění druhého příkladu bylo o dost obtížnější než řešení prvního příkladu. Příčinou bylo především to, že šifrový text v češtině nenahrazoval mezery mezi slovy žádným symbolem, narozdíl od prvního příkladu. Proto jsme nemohli rychle poznat délky jednotlivých slov a museli mnohem více pracovat s frekvencemi jednotlivých bigramů a rozložením samohlásek v českém textu.

Při jednoduché záměně není nutné nahrazovat písmena opět písmeny. Můžeme použít libovolné znaky. Důležité je pouze to, aby různým písmenům a znakům v otevřeném textu odpovídaly různé znaky v šifrovém textu. Zkuste si vyřešit následující příklad šifrového textu. Byla použita jednoduchá záměna, otevřený text je v angličtině a mezery v něm byly vynechány. Pokud se vám řešení nezdaří, můžete si je přečíst v povídce *Zlatý brovek*, jejímž autorem je *Edgar Allan Poe*. Anglický název povídky je *The Gold-Bug*.

5 3 † † † 3 0 5)) 6 * ; 4 8 2 6) 4 † .) 4 †) ; 8 0 6 *
; 4 8 † 8 ¶ 6 0)) 8 5 ; 1 † (; : † * 8 † 8 3 (8 8) 5 *
† ; 4 6 (; 8 8 * 9 6 * ? ; 8) * † (; 4 8 5) ; 5 * † 2 :
* † (; 4 9 5 6 * 2 (5 * - 4) 8 ¶ 8 * ; 4 0 6 9 2 8 5) ;
) 6 † 8) 4 † † ; 1 († 9 ; 4 8 0 8 1 ; 8 : 8 † 1 ; 4 8 † 8
5 ; 4) 4 8 5 † 5 2 8 8 0 6 * 8 1 († 9 ; 4 8 ; (8 8 ; 4 (†
? 3 4 ; 4 8) 4 † ; 1 6 1 ; : 1 8 8 ; † ? ;

Frekvence souhláskových bigramů v českých textech

Při řešení druhého příkladu jsme několikrát použili četnost souhláskových bigramů v českých textech. V následující tabulce je přehled těch nejčastějších. Tabulka byla vytvořena na základě různých českých textů o celkové délce 82 775 písmen.

Bigram	Frekvence
PR	1,98
ST	1,81
CH	1,01
SK	0,63
DN	0,57
SL	0,56
TR	0,46
KT	0,37
TN	0,32
ZN	0,31
SP	0,30
NS	0,26

Cvičení 1.1 *Následující šifrový text byl vytvořen jednoduchou záměnou z českého textu bez mezer. Zkuste najít původní otevřený text.*

HMGPY GEXVX OBYOK UKFUF XILTB PKIYK NIGCU KCKOH LFBPU NGJFK
 HPJEX HLHXP BFNKF GYOKU KFXVK OBINK VKOGH KOZKU PXSJI KYIXH
 IMOKK CKRBO KUGTB MGKYG EFBVR LOGNK YPKNG CXFGM OXCKH BNBOK
 UGIKO UKCXP XYIGF GNXVK YBFKY GEFBV HPBVR LOKYG OBTXX NKNBR
 LOKZX ENKOK MXHUG IKOUL HXEOX EIJMX UKCEK ZBNKK MOKHN XIJVV
 OKUKF PFKIB VCXCO GJFOK OXUGI KOULR LOLEX NNXTX IPFHX RLOGM
 GENXV JCPUL TMUFX IBCHX EKOBO GUXFU ELCHT MIPFO BEGPX RXYKN
 KUKKO XVKOG UEGFG JVXOF KURIK HJINX ZKUGT JYXIK FXETG VJPZX
 NBUXV GFINJ OGPXE KOYGE FBVYO KUKFX VTBVE KOFBV TKPFX ZBKTB
 VTKPF XZBFB VEXOY NXFIH KOGEO GJMGK YBOZK UGFXN UIKFN XCYGF
 UKOTM KPNBU KHCKY IKPXN XVSIX TUJQQ