Прикладная лингвистика

УДК 81'322.2 + 003.26 + 004.415.24 **DOI: 10.14529/ling150410**
ББК Ш100.3 + Ш11

ЛИНГВИСТИЧЕСКАЯ СТЕГАНОГРАФИЯ: СОВРЕМЕННЫЕ ПОДХОДЫ. ЧАСТЬ 2

О.И. Бабина
Южно-Уральский государственный университет, г. Челябинск

Рассмотрены подходы к лингвистической стеганографии, которые направлены на преодоление недостатков формального подхода. В отличие от формальных методов, семантический и трансформационный подходы имеют целью скрыть факт передачи тайного сообщения не только от автоматических систем, но и сделать передачу надежной при анализе стеготекста человеком. Рассмотрены способы учета семантики лингвистических единиц текста. Отдельно выделен подход, основанный на кодировании информации с использованием систем машинного перевода. Выявлены достоинства и недостатки использования описанных методов.

Ключевые слова: лингвистическая стеганография, стеготекст, сокрытие информации, защита информации, автоматическая обработка текстов

Введение

Лингвистическая стеганография является довольно молодой областью и включает в себя разработку методов и способов сокрытия тайного сообщения в тексте на естественном языке. Задача таких методов заключается в построении стеготекста, который в определенном смысле «выглядит» как некоторый обычный текст на естественном языке (то есть непосвященному кажется, что этот текст создан с целью передачи сообщения, складывающегося из смысла лингвистических единиц этого текста). Однако в действительности прагматической функцией такого текста является передача стеганографически скрытого в нем сообщения, не вытекающего из смысла составляющих его лингвистических единиц.

Ранее мы рассматривали формальные подходы к генерации стеготекста, и вывели их существенный недостаток: при формальном подходе не учитывается смысл генерируемого сообщения, поэтому при анализе такого текста человек может заподозрить наличие скрытого сообщения в этом тексте, и цель стеганографии (скрыть факт передачи некоторого сообщения) не будет достигнута. В данной статье мы рассмотрим подходы, направленные на преодоление этого недостатка.

Лексико-семантический подход

Сокрытие информации в тексте на лексико-семантическом уровне включает использование лексических подстановок: в некоторых позициях в предложении одни слова могут быть замещены другими, в определенном смысле, эквивалентными. Такие методы основаны на использовании базы классов эквивалентности: классы представлены набором слов (фраз), каждому из которых произвольно присвоен n-битный номер в двоичной форме (n зависит от мощности множества элементов класса, которая должна быть 2^n). В процессе генерации стеготекста T система замещения просматривает текст-контейнер C слово за словом и выявляет для каждого словоупотребления его класс эквивалентности. Если слово не присутствует ни в одном из классов, оно в неизменном виде копируется в стеготекст T и при этом не несет в себе скрытой информации. Если слово присутствует в каком-либо классе эквивалентности, то система вместо исходного слова текста-контейнера C вставляет слово из его класса эквивалентности с номером, соответствующим n следующим битам тайного сообщения M.

При разработке таких систем усилие сосредоточены на следующих вопросах: Как интерпретировать классы эквивалентности? Как автоматизировать построение классов эквивалентности? Какую процедуру следует использовать при построении стеготекста с применением классов эквивалентности?

Классы экивалентности могут, в частности, отражать формальную вариативность лексических единиц. Так, в [16] в качестве классов эквивалентности выделяются слова на английском языке с американским и британским спеллингом соответственно, которые кодируют 1 бит информации (в каждом таком множестве – 2 элемента). Тогда при составлении стеготекста выбор из пар *favorite – favourite, criticize – criticise, theater – theatre* и т.п. осуществляется в зависимости от значений битов скрываемой информации.

Однако наиболее популярным в лингвистической стеганографии подстановочным методом является метод синонимических замен, где в качест-

ве классов эквивалентности выступают синонимические ряды [7–9, 14, 17, 19.]. Слова-синонимы характеризуются полным или частичным совпадением значений и взаимозаменяемостью в некоторых контекстах, чем обусловлена возможность синонимических замен, в результате которой создается стеготекст, претендующий на сохранение свойств «естественного» текста на уровне лексической семантики. Например, набор предложений

Челябинск – *восхитительный* город
Челябинск – *замечательный* город
Челябинск – *незабываемый* город
Челябинск – *отличный* город

имеет приблизительно одинаковый смысл, и ни одно из предложений не выглядит неестественным не только для машинного анализа, но и для человека-анализатора. Таким образом, любой член синонимического ряда {*восхитительный, замечательный, незабываемый, отличный*} может замещать пробел в рамке «*Челябинск* – _____ *город*». Тогда, присвоив каждому члену этого ряда номер в двоичном коде (например, *восхитительный* – 00*, замечательный* – 01*, незабываемый* – 10*, отличный* – 11), возможно формировать стеготекст *T* путем подстановки в некоторые позиции текста-контейнера *C* соответствующего синонима – такого, номер которого совпадает с текущей последовательностью бит кодируемого сообщения *M*.

Построение классов синонимической эквивалентности представляется довольно трудоемкой задачей. Для автоматизации этой задачи для английского языка зачастую прибегают к информационно-электронному ресурсу WordNet [13], который включает около 70 тыс. слов, объединенных в сеть посредством семантических парадигматических отношений (синонимия, метонимия и т. д.). Вхождением в тезаурусе WordNet является синсет – синонимический ряд. Таким образом, при решении задач лингвистической стеганографии список синсетов может использоваться как набор классов эквивалентности [6–8, 12, 15, 18]. Однако в [18] отмечается, что не все слова могут быть включены в классы эквивалентности: многозначные слова могут входить в различные синсеты, и в общем случае, в различных синсетах таким словам могут быть приписаны разные коды. Использование такого слова при кодировании приведет к неоднозначной интерпретации сообщения, заложенного в стеготексте. Поэтому в классы эквивалентности должны войти не все элементы, а лишь удовлетворяющие определенным критериям отбора. В [18] предлагается формировать классы эквивалентности только из однословных единиц, которые входят лишь в один синсет, или – если один синсет является полностью вложенным в другой – использовать меньший по объему синсет в качестве одного из классов эквивалентности. Экспериментально установлено, что около 30 % слов попадут в классы эквивалентности, сформированные в соответствии с этими правилами, что позволяет построить довольно внушительный словарь для синонимических замен.

Другое направление решения проблемы множественной принадлежности синсетам – применение автоматических методов различения смыслов слов (word sense disambiguation), позволяющих идентифицировать значение, в котором используется слово, что даст возможность отнести его к правильному синсету. В пределах синсета – код каждого слова уникален.

Для формирования классов синонимической эквивалентности в [2] указывается на возможность использования антонимических пар – в каждое множество синонимических замен на русском языке может входить лексема и морфологически производное ее антонима с приставкой *не-*: *легко – нетрудно, большой – немаленький* и т.п. При наличии машиночитаемого словаря антонимов, процедура построения таких классов эквивалентности может показаться тривиальной: 1) для исходного слова без приставки *не-* в своем составе, добавление приставки *не-* к его антонимам (без приставки *не-*) из словаря; 2) для исходного слова с приставкой *не-* в своем составе, формирование его производного путем удаления приставки *не-* и далее поиск по словарю антонимов (без приставки *не–*) к этому производному. Однако в обоих случаях требуется обязательная проверка по словнику на предмет существования в языке «искусственно» синтезированной формы с или без приставки *не-* соответственно, так как в языке, в частности, существуют слова, для которых утеряны исходные формы без *не-* (например, *непобедимый, ненароком, негаданный, невредимый, невпопад* и т.п.), значит, формы **победимый, *нароком* и т.п. не могут использоваться в подстановках. При этом остается проблема многозначности: в словосочетаниях *скользкая* (*дорога*) – *скользкий* (*начальник*), *крепкий* (*чай*) – *крепкая* (*стена*) и т.п. прилагательные используются в своих разных значениях, для каждого из которых имеется свой антоним. Возникает проблема использования нужного антонима для формирования производного с *не-* при построении классов эквивалентности и реализации подстановок. Кроме того, словообразование с помощью *не-* не всегда дает производное слово с противоположным значением – семантика сохранившихся в языке форм с и без *не-* может значительно отличаться, и тогда добавление *не-* к антониму исходного слова не приводит к формированию его синонима, например, *бессодержательный* (пример) [исходное слово] → *наглядный* (пример) [антоним исходного слова] →**ненаглядный* (пример) [не синоним исходного слова].

Построение стеготекста с применением классов эквивалентности может заключаться в тривиальной автоматической замене одного синонима на другой (с кодом, соответствующим битам тай-

ного сообщения) каждый раз, когда текущее слово текста-контейнера попадает в один из классов эквивалентности. Проблема с применением такого подхода состоит в том, что синонимы являются лишь частично взаимозаменяемыми: в отдельных случаях синонимическая замена может приводить к нарушению лексической сочетаемости (например, *сделать вывод* – **выполнить вывод*). Это приводит к мысли, что синонимические замены следует проводить с учетом контекста, в котором употребляется слово в тексте-контейнере. Для учета этого фактора в процедуру внедрения тайного сообщения в текст-контейнер вводится этап проверки результирующего предложения по *n*-грамам [5; 8]. Таким образом, классы эквивалентности определяются *динамически* в процессе формирования стеготекста, и включают лишь те лексические единицы, которые сочетаются (с достаточно высоким частотным показателем) с контекстными для данного слова лексическими единицами из текста-контейнера. Например, в предложении «*Пять подземных толчков зарегистрировано за сутки*» на основе предиката *зарегистрированный* может быть построен синонимический ряд {*зарегистрированный, зафиксированный, замеченный, отмеченный, закрепленный, прикрепленный, помеченный, выделенный, упомянутый, отпразднованный*}. Однако проверка по *n*-грамам покажет, что с единицей *подземные толчки* лексически сочетаются лишь первые четыре элемента этого ряда. Именно они и должны сформировать динамический класс эквивалентности для данного случая.

Для получения после подстановок грамматически верных предложений требуется учет морфологии, особенно в высоко флективных языках. Эта проблема обычно решается использованием морфологических анализаторов или эвристиками, как правило, основанными на анализе окончаний слов из текста-контейнера, подлежащих замене.

При моделировании синонимических замен одной из задач является сохранение характеристик статистического распределения слов, используемых в качестве заместителей. В этой области, наряду с алгоритмами сжатия (например, в [15], для соотнесения битов тайного сообщения со словами-субститутами применяется кодирование Хафмана), применяется стохастическое моделирование. Так, в [1] для сохранения частотного распределения синонимов в стеготексте предлагается применить процедуру, соответствующую методу Монте-Карло, где каждая синонимическая замена рассматривается как одна реализация случайного процесса по выбору синонима из класса эквивалентности: выбор слова-субститута зависит от значения случайной величины, полученного при очередной реализации стохастического процесса, и текущей последовательности бит тайного сообщения. Такой подход приводит к тому, что, в общем случае, при шифровании одной и той же последовательности бит тайного сообщения в стеготексте могут использоваться разные слова-субституты из одного и того же класса эквивалентности. Тем самым достигается эффект рассеивания и перемешивания (принцип Шеннона), обеспечивающий определенную криптографическую стойкость. Для расшифровки сообщения потребуется синхронизация двух стохастических процессов (на стороне отправителя и получателя сообщения), то есть сгенерированная последовательность значений случайной величины выступает в роли ключа.

Недостатком использования подхода синонимических подстановок является то, что в качестве текста-контейнера выступает некоторый ранее созданный текст, обычно, имеющийся в публичном доступе. Простое сравнение двух текстов позволит легко обнаружить наличие измененных компонентов. Как способ преодоления этого недостатка в [10] предлагается отойти от идеи автоматизации внедрения тайного сообщения в существующий текст-контейнер: текст, включающий зашифрованное сообщение, должен быть построен *вручную* на основе списка кодирующих тайное сообщение слов, сгенерированного автоматически. Классы эквивалентности используются для автоматической генерации такого списка. Имея список, человек должен достроить «шумовой» текст, используя слова, не входящие ни в один из классов эквивалентности (проверка может быть автоматизирована с использованием специального текстового редактора), и не изменяя порядка следования автоматически сгенерированного списка слов.

Онтологический подход

Использование онтологического подхода подразумевает использование эксплицитной модели знаний о мире, которые могут вариативно быть выражены в языке. Как и в предыдущем подходе, скрытую информацию несет выбор из альтернативных способов языкового выражения значения, при этом задача состоит в том, чтобы при внедрении тайного сообщения сгенерированный текст *T* был семантическим эквивалентен исходному тексту-контейнеру *C*. Способ применения такого подхода показан в [3; 4]. Использование онтологического подхода подразумевает использование в качестве статической базы знаний онтологии, лексикона и представлений Смысл – Текст. Динамический компонент представлен парсером, анализатором и генератором.

Посредством аппарата ролевой семантики в онтологии значения концептов представлены с помощью множества семантических признаков, а также фреймов вида «атрибут: значение», где атрибутами являются ролевые отношения участников описываемой концептом ситуации, и на значение атрибутов могут быть наложены семантические ограничения. В ходе анализа каждое предложение текста на естественном языке репрезентируется посредством представления Смысл–

Текст: ролевые структуры онтологии заполняются конкретными языковыми выражениями и таким образом формируют дерево зависимостей с типизированными семантическими отношениями между участниками в описываемой ситуации. Например, для предложения *Маша сказала Ивану, что она едет в Москву* представление Смысл-Текст примет вид дерева, представленного на рисунке.

Представление Смысл – Текст

С использованием секретного ключа, каждое предложение s_k текста может быть преобразовано к битовой строке B_k, которая используется для маркировки предложений, скрывающих информацию, и для собственно представления битов скрытого сообщения *M*. В общем случае B_k маркированного предложения в исходном тексте-контейнере *C* может не совпадать с текущей последовательностью из тайного сообщения *M*. Поэтому предложение s_k должно подвергнуться различным трансформациям, пока не будет найдено его преобразование s'_k, которое при применении секретного ключа дает B_k, точно совпадающее с текущей последовательностью бит из сообщения *M*, которую необходимо зашифровать. Преобразованное таким образом предложение s'_k включается в стеготекст *T* вместо исходного s_k.

Преобразования, которые производятся в процессе подбора подходящего B_k, не должны в значительной мере менять семантическое наполнение текста. Такие трансформации могут представлять собой синтаксический преобразования [3], и в этом случае не обязательно построение представлений Смысл – Текст – достаточно построить дерево синтаксического разбора предложения с использованием какого-либо парсера. Однако в [4] в качестве трансформаций, сохраняющих смысл текста, также выделяются три более сложных типа, требующих обязательного анализа семантической структуры трансформируемого предложения:

– *расширение*: вставка информации из другого предложения;

– *сокращение*: удаление повторяющейся информации;

– *подстановка*: изменение эквивалентной информации.

Первые два типа трансформаций можно осуществлять на основе установленных кореферентных связей в тексте, объединяя деревья представлений Смысл – Текст двух предложений (расширение) или удаляя семантически необязательные ветви дерева, если в них содержится информация, присутствующая также в других частях текста (сокращение). Последний тип трансформаций осуществим путем вставки информации из составленной заранее базы фактов (онтологии).

Использование онтологического подхода предоставляет схему внедрения тайного сообщения, стойкую для различных преобразований над текстом *T* (которым он может подвергнуться с целью не допустить того, чтобы сообщение достигло адресата): синонимические подстановки, синтаксические преобразования (например, пассивизация), перевод на другой язык не меняют семантической структуры предложения. Однако автоматическое распознавание этой структуры – отдельная, наиболее трудно поддающаяся формальному моделированию задача. Кроме того, доступность развернутых онтологий, содержащих достаточную для реальных задач базу знаний, весьма ограничена. Несмотря на множество доступных программных ресурсов для построения онтологий, разработанные формализмы для представления знаний о фактах действительности, реальные базы знаний о мире с достаточным количеством информации для применения в реальных приложениях практически отсутствуют. В результате реализация такого подхода требует значительной предварительной работы по созданию всех компонентов базы лингвистических знаний.

Подход, основанный на машинном переводе

Отдельно следует выделить подход, основанный на применении систем машинного перевода [11]. Данный подход в определенной степени противоположен предыдущим, основная задача которых состояла в определения способа построения стеготекста *T*, не отличимого от текста, созданного человеком для передачи только того смысла, который может быть извлечен из значений составляющих его лингвистических единиц. Перевод текста на другой язык можно рассматривать как межъязыковой парафраз: при переводе возможны различные способы выражения одной и той же мысли. Альтернативность переводов предоставляет возможность выбора, а значит, инструмент для внедрения тайного сообщения: для внесения сообщения *M* в текст могут использоваться переводы, выполненные несколькими машинными переводчиками, и выбор одного из таких переводов в каждом случае кодирует определенное количество информации. Очевидно, носителями информации здесь выступают такие предложения, которые содержат проблемные (с точки зрения машинного перевода) языковые выражения, задача перевода которых решается различным способом в различных системах перевода.

Кроме того, машинный перевод текста может подвергнуться пост-редактированию, в ходе которого в перевод добавляются (или исправляются) ошибки, типичные для систем машинного перевода, таких как неправильный перевод или пропуск функциональных слов, неверное согласование подлежащего и сказуемого, пословный перевод, неверный выбор лексического эквивалента, транслитерация единиц, отсутствующих в словаре, перевод имен собственных, формально совпадающих с нарицательными (*Надежда – Норе*), неверная видо-временная форма глагола (что обусловлено различиями в составе грамматических категорий глаголов в различных языках) и другие. Идея данной реализации подхода заключается в том, чтобы замаскировать биты скрытого сообщения под «шумы» (ошибки), получаемые в результате машинного перевода. Наличие ошибки определенного типа несет в себе информацию о содержании тайного сообщения.

Применимость такого подхода обусловлена тем, что в современных условиях чтение текстов, полученных в результате машинного перевода, является нормой: большие потоки данных легче обрабатываются, если они представлены на родном языке, в этом случае имеется возможность быстро ознакомиться с содержанием текста и принять решение о его полезности. Поэтому, несмотря на то, что получаемый при таком подходе стеготекст не вполне соответствует понятию «естественности» – идеалу, который пытаются достичь в предыдущих подходах, тем не менее ситуация чтения такого текста, заведомо содержащего ошибки, «естественна» в своем роде.

Заключение

Анализ современных лингвистических методов стеганографии позволяет сделать вывод о том, что они основаны на разработке достаточно большой базы лингвистических знаний, которая в ряде подходов, фактически, выступает ключом для расшифровки сообщений, внедренных в стеготекст. Как и для задач криптографии, остается проблема передачи такого «ключа» получателю сообщения – такая передача должна осуществляться по незащищенному каналу. Отчасти эту проблему может решить за счет использования доступных для общего пользования источников – текстов-контейнеров, доступных через Интернет, лингвистических ресурсов (например, WordNet, парсеры и т. д.).

Для решения этой проблемы необходима разработка таких методов генерации стеготекста, интерпретация которого зависела бы от относительно короткого ключа (например, простого числа). Такие работы проводятся в двух направлениях: во-первых, разработка процедуры преобразования текста-контейнера в стеготекст, однозначно определяющейся ключом. Во-вторых, использование в качестве тайного сообщения потока бит, полученного в результате применения одной из схем шифрования.

Из рассмотренных подходов наиболее стойким к стегоатакам и, вместе с тем, наиболее сложным для реализации представляется онтологический подход. Общей тенденцией лингвистической стеганографии является усложнение лингвистической базы знаний с целью обеспечения большей надежности сокрытия информации в тексте. В связи с этим, развитие методов лингвистической стеганографии неразрывно связано с достижением определенных высот в области автоматической обработки текстов.

Литература/References

1. Алиев А.Т. Лингвистическая стеганография на основе замены синонимов для текстов на русском языке. Известия Южного федерального университета. Технические науки. 2010. № 11(112). С. 162–171. [Aliev A.T. (Linguistic Steganography Based on Synonym Substitution for Texts in Russian), *The Bulletin of the South Federal University, Technical Sciences*, 2010, no. 11(112), pp. 162–171 (in Russ.)].

2. Ефременко Н.В. Лингвистическая стеганография. Вестник Московского государственного лингвистического университета. 2011. № 619. С. 66–73. [Efremenko N.V. (Linguistic Steganography). *The Bulletin of the Moscow State Linguistic University*, 2011, no. 619, pp. 66–73. (in Russ.)]

3. Atallah M. J., Raskin V., Crogan M., Hempelmann C., Kerschbaum F., Mohamed D., and Naik S. Natural Language Watermarking: Design, Analysis, and a Proof-of-Concept Implementation, I.S. Moskowitz (ed.), *Information Hiding: Fourth International Workshop*, Lecture Notes in Computer Science 2137, Springer, April 2001, pp. 185–199.

4. Atallah M.J., Raskin Viktor, Hempelmann Christian F., Karahan Mercan, Sion Rodu, Topkara Umut, Triezenberg Katrina E. Natural Language Watermarking and Tamperproofing, *Information Hiding: 5th International Workshop* (Noordwijkerhout, The Netherlands), Springer, October 2002, pp. 196–212.

5. Bolshakov I.A. A Method of Linguistic Steganography Based on Collocationally-Verified Synonym, *Information Hiding: 6th International Workshop*, Vol. 3200, Toronto, Canada, 2004, pp. 180–191.

6. Bolshakov I.A., Gelbukh A. Synonymous Paraphrasing Using WordNet and Internet, F. Meziane, E. Metais (eds.), *Natural Language Processing and Information Systems, Lecture Notes in Computer Science*, Vol. 3136, Springer, 2004, pp. 312–323.

7. Chang Ching-Yun, Clark Stephen. Practical Linguistic Steganography Using Contextual Synonym Substitution and Vertex Colour Coding, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (EMNLP ’10), Cambridge, MA, 2010, pp. 1194–1203.

8. Chang Ching-Yun, Clark Stephen. Practical Linguistic Steganography using Contextual Synonym

Substitution and Novel Vertex Coding Method, *Computational Linguistics*, Vol. 40, No. 2, June 2014, pp. 403–448.

9. Chapman Mark, Davida George I., Marc Rennhard. A Practical and Effective Approach to Large-Scale Automated Linguistic Steganography, David G.I., Frankel Y. (eds.), *Information Security: 4th International Conference, ISC 2001*, *Lecture Notes in Computer Science,* Vol. 2200, Springer-Verlag, Berlin, Germany, 2001, pp. 156–165.

10. Grosvald Michael, Orgun C. Orhan. Free from the Cover Text: A Human-Generated Natural Language Approach to Text-Based Steganography, Journal of Information Hiding and Multimedia Signal Processing, vol. 2, no. 2, April 2011, pp. 133–141.

11. Grothoff Christian, Grothoff Krista, Alkhutova Ludmila, Stutsman Ryan, Atallah Mikhail J. Translation-Based Steganography, *Journal of Computer Security,* vol. 17, no. 3, August 2009, pp. 269–303.

12. Meral Hasan Mesut, Sankur Bülent, Özsoy A. Sumru, Güngör Tunga, Sevinç Emre. Natural Language Watermarking via Morphosyntactic Alterations, Computer Speech and Language, vol. 23, 2009, pp. 107–125.

13. Miller George A., Beckwith Richard, Fellbaum Christiane, Gross Derek, Miller Katherine, Tengi Randee. *Five Papers on WordNet*, Technical Report, Cognitive Science Laboratory, Princeton University, 1993. Available at: http://wordnetcode.princeton.edu/ 5papers.pdf (access date: Nov 02, 2014)

14. Morran Michael, Weir George R.S. An Approach to Textual Steganography, S. Tenreiro de Magalhaes, H. Jahankhani, A.G. Hessami (eds.), *Global Security, Safety and Sustainability: Proceedings of the 6th International Conference, ICGS3 2010*, *Communications in Computer and Information Science,* Vol. 92, Braga, Portugal, September 1–3, 2010, pp. 48–54.

15. Nanhe Aniket M., Mayuresh P. Kunjir, Sumedh V. Sakdeo. Improved Synonym Approach to Linguistic Steganography Design and Proof-of-Concept Implementation. 2008. Available at: http://dsl.serc.iisc.ernet.in/~mayuresh/FImprovedSynonymAp proachToLinguisticSteganography.pdf

16. Shirali-Shahreza M. Text Steganography by Changing Words Spelling, *Proceedings of the 10th International Conference on Advanced Communication Technology* (February 17–20, 2008), Vol. 3, 2008, pp. 1912–1913.

17. Topkara Umut, Topkara Mercan, Atallah Mikhail J. The Hiding Virtues of Ambiguity: Quantifiably Resilient Watermarking of Natural Language Text through Synonym Substitutions, Proceedings of the 8th Workshop on Multimedia and Security (MM&Sec'06), Geneva, Switzerland, September 26–27, 2006, pp. 164–174.

18. Winstein Keith. Lexical Steganography Through Adaptive Modulation of the Word Choice Hash, 1998. Available at: http://web.mit.edu/keithw/tlex/ (access date: Nov 02, 2014).

19. Wysuer Brecht, Wouters Karel, Preneel Bart. Lexical Natural Language Steganography Systems with Human Interaction, *Proceedings of the 6th European Conference on Information Warefare and Security* (Shrivenham, UK, July 2-3, 2007), 2007, pp. 303–312.

Бабина Ольга Ивановна, кандидат филологических наук, доцент, доцент кафедры лингвистики и межкультурной коммуникации, Южно-Уральский государственный университет (Челябинск), babinaoi@susu.ac.ru

Поступила в редакцию 4 июня 2015 г.

DOI: 10.14529/ling150410

LINGUISTIC STEGANOGRAPHY: STATE-OF-THE-ART. PART 2

O.I. Babina, *babinaoi@susu.ac.ru*
South Ural State University, Chelyabinsk, Russian Federation

In the article the approaches aimed at overcoming the drawbacks of the formal approach are presented. Unlike the formal methods, semantic and transformational approaches are directed at concealing the fact of passing the message not only from automatic systems, but also from being revealed by a human. Different methods accounting for the semantics of the generated stegotext are observed. Machine translation based steganography is described as a special type of linguistic steganography. Advantages and disadvantages of the techniques presented are analyzed.

Keywords: linguistic steganography, stegotext, information hiding, information security, natural language processing.

Olga I. Babina, Candidate of Philology (PhD), Associate Professor, Associate Professor of the Department of Linguistics and Intercultural Communication, South Ural State University (Chelyabinsk), babinaoi@susu.ac.ru

Received 4 June 2015

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Бабина, О.И. Лингвистическая стеганография: современные подходы. Часть 2 / О.И. Бабина // Вестник ЮУрГУ. Серия «Лингвистика». – 2015. – Т. 12, № 4. – С. 49–55. DOI: 10.14529/ling150410

FOR CITATION

Babina O.I. Linguistic Steganography: State-of-the-Art. Part 2. *Bulletin of the South Ural State University. Ser. Linguistics*. 2015, vol. 12, no. 4, pp. 49–55. (in Russ.). DOI: 10.14529/ling150410