

Speakin' Mumbo-Jumbo

The Dangers of Copying & Pasting



[Ctrl+C] & [Ctrl+V] = trouble



Table of Contents

Introduction

The A,B,C's: Understanding the Roots of the Problem

- A short primer on ASCII for Marketers
- Encoding in a Nutshell
- Non-ASCII by Mistake ... a How-To
- Where the problems occur – 3 hot spots

From A,B,C to Mumbo-Jumbo: Study Results

- Test methodology
- Results: Broken at 41% to 53% of ISPs
- Close-up of broken subject lines and brand impact
- Other critical findings, observations & limitations

Back to Class: Conclusions & Tips

- Conclusion
- 5 quick tips

This Report is For:

- Designers/
Copywriters
- Email Marketers
- ESP/Agency Account
Managers
- Technical Staff at
Email Deployment
Vendors

Introduction

Ever seen an email where the subject line looks more like hieroglyphs or modern art than a legible typography? Through our deliverability solutions, we review hundreds of emails daily from the world's top brands across a broad range of industries. We see great subject lines but, unfortunately, we also see many "broken" ones such as these:

You've Won, today's your lucky day! =?UNKNOWN?B?lg==?= Final 3 Days!
You've Won, today's your lucky day! Å Final 3 Days!
You've Won, today's your lucky day! □ Final 3 Days!

Curiosity led us to ask “**what’s causing subject lines to look so ‘broken’?**” After an initial round of investigation, we identified an alarming trend in the production of email: the use of word processors like Microsoft’s Word to create HTML email. Unfortunately, as this report will show, the effects of copying characters & code from Microsoft Word into a deployment system can be devastating to your brand.

Once we identified a potential source for the broken characters, we did more investigation to better understand how language is transferred from one computer to another and then rendered. This brought us into close quarters with two concepts, ASCII and Encoding, that are often kicked around the industry as standard nomenclature but rarely understood or treated with enough respect.

We’ve learned that those who don’t respect the limitations of ASCII and the importance of choosing the right encoding, risk losing their well thought-out subject lines in a storm of characters and glyphs that can’t be understood by the most astute linguist much less the rest of us mere mortals.

This report covers our findings and provides practical, understandable advice on how you can avoid speakin’ mumbo jumbo to your customers.

The A, B, C's Understanding the root of the problem



A short primer on ASCII for Marketers

What is ASCII?

When you type “Hello World” on a screen, a computer has to convert each letter into a language that it, the computer, can understand. The translation between machine language and a Latin based alphabet humans can read happens through ASCII (American Standard Code for Information Interchange). The ASCII standard, born in 1960 and under development until 1986, is the intermediary language between computer-based binary, ones and zeros, and Latin characters or glyphs as they’re sometimes called.

ASCII stores letters as numbers that can be read and interpreted by a computer. The standard ASCII character set consists of a total of 128 characters; 32 non-printable control characters and 96 printable-characters (which include punctuation, numbers, and the 26 letters of the English alphabet):

```
!"#$%&'()*+,-./  
0123456789:;<=>?  
@ABCDEFGHIJKLMNO  
PQRSTUVWXYZ[\]^_  
`abcdefghijklmnop  
pqrstuvwxyz{|}~
```

The visible ASCII Character Set

95 visible characters of ASCII +
Space for 96 ‘printable’ characters
total.

Why is it so important?

Simple. Unless you specify an encoding (we’ll get to that scary term in a second), the receiving computer will assume your subject line is in ASCII. If it’s not and no encoding has been provided, the computer will display it to the best of its ability (which yields an “A” for effort but, as you’ll see, an “F” in performance).

Encoding In A Nutshell

Computers assume the characters are in ASCII unless you tell them otherwise. This is where “Encoding” comes in; it is how you tell a computer which language your message is in (think of it as that ‘translator’ you wish you had with you on your last trip to Timbuktu). The computer takes the encoding “character set” and uses it to display the message to the screen in the proper language with the proper punctuation. To ensure that it works right, the sending side has to properly encode it and the receiving side has to have the ability to render the specified character set. Neither happens consistently or universally!

A Mailer creates a German-language email. Since the email contains non-ASCII characters, such as umlauts, the mailer is smart enough to Encode it using the Western European Character-Set.

Email Sent

The email is received by the ISP/Email-Reader and the character encoding is interpreted

HAPPY FACTOID
Over 1000 different encodings exist!

SAD FACTOID
ISPs/Email-Readers differ in their ability to interpret them, with none being able to interpret them all!

If the ISP can read and interpret the encoding, then the message will mostly likely appear as was intended.

If the ISP can't read the encoding character set then the message will appear broken with supplanted characters in place of ones that can't be read and interpreted.

Non-ASCII.. (by Mistake) .. in Microsoft

Non-ASCII characters are common-place. If you write your emails in any foreign language, you already know this (or should!) and are encoding your email subject-lines and content accordingly. Most of us non-techies don't know what ASCII is but where many of even the most savvy-international marketers get in trouble is when they assume what looks like ASCII is ASCII. And ..maybe it would be ... if it wasn't for that omni-present Microsoft and their special way of doing things. Let us show you what we mean..

Subject line typed in an ASCII-compliant editor:

"You've Won - today's your lucky day!"

Typed into Microsoft Word or the body of an Outlook email:

“You’ve Won – today’s your lucky day!”

Copy & paste the subject line typed in Word into an ASCII reader and this is what you see:

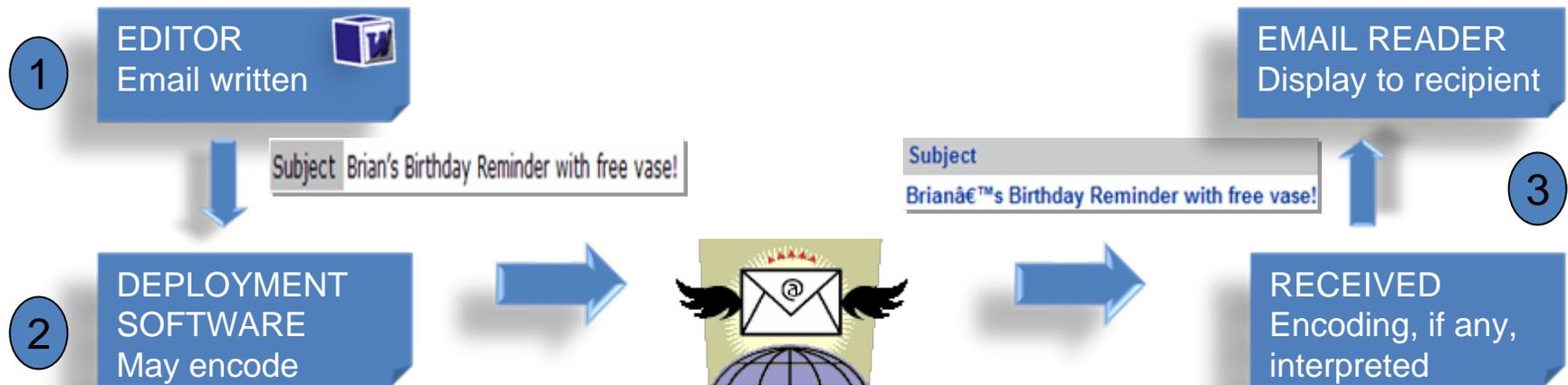
!You've Won ! today's your lucky day!!

Notice the quotes & apostrophes are straight in the ASCII editor but are curled when typed in Word.

What's happening in the example above?

- Word defaults to using its own special **curly** quotes and apostrophes . These are non-ASCII characters; when pasted into an ASCII reader/ editor, they can “break” as seen in the 3rd example.
- In addition to the non-ASCII quotes, the dash is not a true dash. When you type a dash in Word with a space on either side of it, Word converts that from a standard dash to an “**en-dash**” which is essentially longer than a normal dash (double length) and...again.. non-ASCII.

Where can the problems occur?... 3 Hot Spots



1. The editing software in which the email is written may permit non-ASCII. This is not a problem per se as non-ASCII characters may be desired such as copyright symbols and foreign characters. The problem is with editors like Word that take traditional ASCII-characters such as an apostrophe and convert them to non-ASCII without the author realizing it.
2. The deployment software does not enforce ASCII subjects or encoding on non-ASCII subjects. Meaning, it permits non-ASCII subject lines to be mailed un-encoded. Note, even if the software did enforce encoding on non-ASCII, data pasted from Word would still cause an issue as there is no known encoding specifically for 'translating' Microsoft's special characters.
3. The receiving software must interpret the subject. It assumes ASCII if no encoding or, if encoded, will interpret the encoding. The problem is that each email reader/receiving software is unique in its ability to display non-ASCII, non-encoded data AND, equally important, can not interpret all encodings properly so even encoded data may display improperly from one email reader to the next.

From A, B, C to Mumbo-Jumbo Test Results



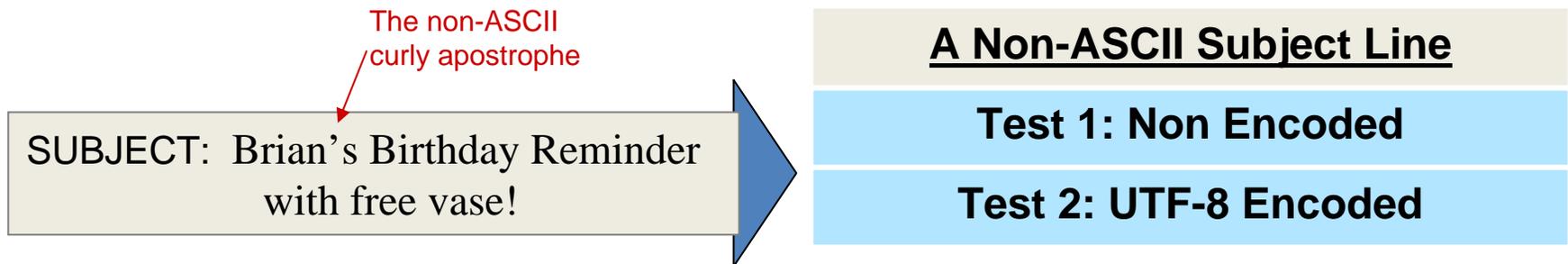
Our Testing Methodology

Objective: Test which ISPs properly display Word's version of Non-ASCII characters in the subject line

We conducted the test by creating two subject lines and sending them out to 18 top ISPs to assess how they would handle non ASCII characters in subject lines and whether encoding would make any difference.

Test 1: The first subject line was written in Microsoft Word and copied and pasted it into our email broadcaster for distribution. We specified no encoding since this is the most common mistake we see (mailer assumes it looks like ASCII or doesn't know what encoding is, so no encoding is applied).

Test 2: The second subject line was created identically to the first. However, this time we encoded the subject line using UTF-8*. Why did we choose UTF-8 encoding? Because its one of the most common encodings and, also, because there is no special "Microsoft character" encoding.



* UTF-8 – Unicode Transformation Format – Variable length character encoding based on ASCII that is becoming the preferred encoding for e-mail and web pages.

The Results: broken at 41-53% of ISPs

- In a few isolated cases, the subject rendered properly in the **List View** (the listing of emails the customer first sees in their mailbox) but improperly in the **Message View** (what they see after they click to open a particular email) and vice versa. These include: Excite, Mail.com, and Verizon.
- Overall, the **non-ASCII, non-encoded subject line broke** / rendered incorrectly (designated by an “X”) in both the List View and Message View at **44%** of the top tier ISPs. Some of the notable ones: AOL, Cablevision, and Earthlink.
- Overall, the **non-ASCII, encoded version** fared just as poorly – showing up broken in the **List View at 41%** of the top tier ISPs and showing up broken in the **Message View at 53%** of the ISPs. Some of the notable ones: Hotmail, Yahoo, and Comcast.

ISP	Non-ASCII, Non-Encoded		Non-ASCII, UTF-8 Encoded	
	List View	Msg View	List View	Msg View
AOL	X	X	✓	✓
AT&T	✓	✓	X	X
Bellsouth	✓	✓	X	X
Cablevision	X	X	ND	ND
Comcast	✓	✓	X	X
CompuServe	X	X	✓	✓
Cox	X	X	✓	✓
Earthlink	X	X	✓	✓
Excite	X	✓	X	X
Gmail	✓	✓	✓	✓
Hotmail/MSN	✓	✓	X	X
Lycos	X	X	✓	✓
Mail.com	✓	✓	✓	X
Netzero	✓	✓	X	X
Road Runner	✓	✓	✓	✓
Usa.net	X	X	✓	✓
Verizon	✓	X	✓	X
Yahoo	✓	✓	X	X

ND = Email was not delivered by the ISP so the results could not be evaluated. Whether the non-delivery is due to the format is unknown.

A close-up look at the “broken” subject lines

- The manner in which the subject line was “broken” differed across ISPs, reflecting the diverse capabilities of email readers and the difficulty in projecting what your customer will see.

Properly Rendered	Road Runner	✓	Subject Brian's Birthday Reminder with free vase!
BROKEN Rendering (Non-Encoded)	AOL	✗	Subject: Brian's Birthday Reminder with free vase! Date: Sat, 21 Apr 2007 3:00 PM
	Netzero	✗	Subject Brian's Birthday Reminder with free vase! Last chance! Start your FREE trial today
	Earthlink	✗	Subject: Brian's Birthday Reminder with free vase! Date: Apr 21, 2007 3:00 PM
Broken Rendering (UTF-8 – Encoded)	AT&T	✗	Subject: Brian's Birthday Reminder with free vase! Date: Sat, 14 Apr 2007 21:28:04 +0000 [View Source]
	Yahoo	✗	Subject Brian's Birthday Reminder with free vase!

Other critical findings, observations, & limitations

- During our research we spoke with both in-house mailers using home-grown and packaged software as well as many of the large ESPs (Email Service Providers). Without exception, none knew why the subject lines weren't rendering properly and many were **not even aware the problem existed** - even though it was directly affecting their own or their clients' campaigns.
- No email software or email service provider we reviewed enforced* ASCII subject lines nor did they enforce a particular encoding even when a list of possible encodings was made available. This means, mailers and ESP full-service personnel are largely **left to their own best-judgment** as to what's ASCII, what's not, and what specific encoding (out of over a thousand), if any, to use.
- The results highlighted in this report reflect how the subject line fared in both the list and message view. The 'list view' is the **first view** your customer sees and is the display of the list of emails (e.g. by date, from, and subject line) in their mailbox and the 'message view' is what they see when they actually **click to open** the message. While this study did not focus on the body of the email, we suspect, given the variable nature of email readers, that how the subject renders may not always match how the body of the message renders. Given the importance of the subject line in compelling the customer to open or discard the email, we believe it merits the special attention we've given it.
- There are **hundreds, if not thousands, of editors** in which a mailer might write their copy. In this study we highlighted issues associated with producing content in Microsoft Word and then pasting that in the email deployment system. Is it probable that there are other tools being used to write emails, that also produce or allow non-ASCII characters that may appear to be ASCII? Of course.

*Deployment software did not prevent non-ASCII data from being copied into the subject line field and/or did not identify the subject as being non-ASCII so the mailer would know to change it to ASCII or encode it.

Back To Class Conclusions & Parting Advice



Conclusion – How do you stack-up?

While we have already alluded to this point, its worth reiterating as its by far the most important point of the study and goes well beyond the specific copying & pasting of subject lines from Word that this particular study dealt with:

Email Readers (e.g. Hotmail, Yahoo, Lotus Notes, Outlook) are not created equal.

How your email renders in one, is not proof of how it will render in all the others. If you QA your email in AOL, you know one thing and one thing only – what your email looks like and what your customers' experience will be in AOL. Unfortunately, you have absolutely no inkling of its rendering in Hotmail or Lotus Notes or Express or Earthlink, etc. Prior studies illustrated this fact in respect to the body of the message – this study clearly illustrates it as it pertains to the subject line.

When we began this project neither we, nor the affected mailers knew why the subject lines were being broken. While we soon solved that mystery, one mystery remains unsolved: despite the known fact that different email readers render emails differently, **why are so many mailers (from the largest companies in the world to the tiniest Mom&Pops) still sending emails that render poorly?** Is it because..

- (As we suspect), most mailers are still only QA'ing their emails in one email reader and thus don't realize the problem? Or
- Mailers understand the issue but do not care?
- Mailers understand and care but don't realize there are solutions available that allow them to view their emails across readers efficiently and cost-effectively?

If you're a mailer, have some fun, take this anonymous [survey](#) & instantly find out how you stack-up vs. others!

Parting Advice - - 5 Quick Tips

1. Do **not copy email subject lines from Microsoft Word** to your deployment system for mailing. Even if your email deployment system manages to convert the non-ASCII characters to ASCII, it's simply not worth the risk. You are better off typing your subject line directly into the email software or ESP's editor rather than pasting it from elsewhere.
2. If your copy is "**English**" (or more specifically, contains only characters in the ASCII character set) and/or you do not use encoding, **ensure the solution you are using to actually write your subject line is an ASCII editor**. As you've seen, Microsoft Word takes typical ASCII characters such as quotes and makes them non-ASCII, so it is not always so obvious.
3. When encoding emails, keep in mind the **subject line is encoded separately from the body of the content**. In fact, they can have two different encodings (e.g. no encoding in the subject vs. a specific encoding in the body). If you need to encode your emails and do not see both a subject line and a body encoding option, speak to your ESP or email software provider as their solution may have limitations that don't allow encoding on one or both.
4. If your subject **contains non-ASCII characters** (e.g. Foreign language characters, copyright symbols), **you should be encoding it!** There are over a thousand different encodings to choose from. Absolutely don't guess – seek guidance. Then remember to test because each email reader differs in respect to which encodings it can successfully interpret and render.
5. **TEST TEST TEST and do so in more than 1 email reader!!!!** Pivotal Veracity clients – use your [eDesign Optimizer](#) solution to easily evaluate what your email looks like in dozens of US, Asian, European, and software email readers to ensure all aspects of your email look and function right.

Thank you!

-Your Team at Pivotal Veracity

www.pivotalveracity.com

3420 E. Shea Blvd. Ste. 120

Phoenix, AZ 85028

602-971-0502

If you have any questions or would like to attend an upcoming encoding webinar please contact us at support@pivotalveracity.com