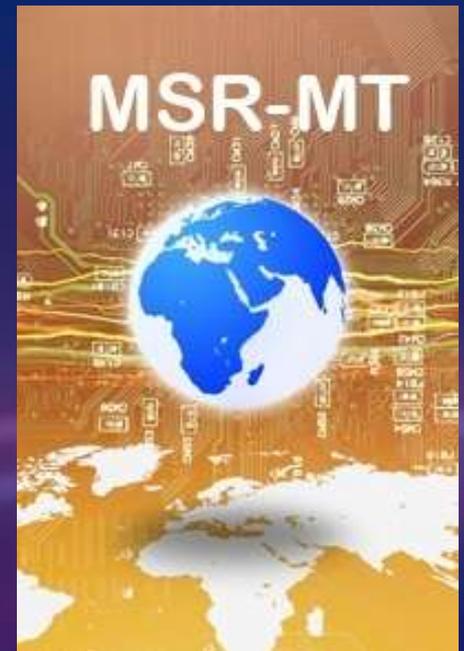


Microsoft Machine Translation: From Research to Real User

MT Summit XI

September 12, 2007



Microsoft

Outline

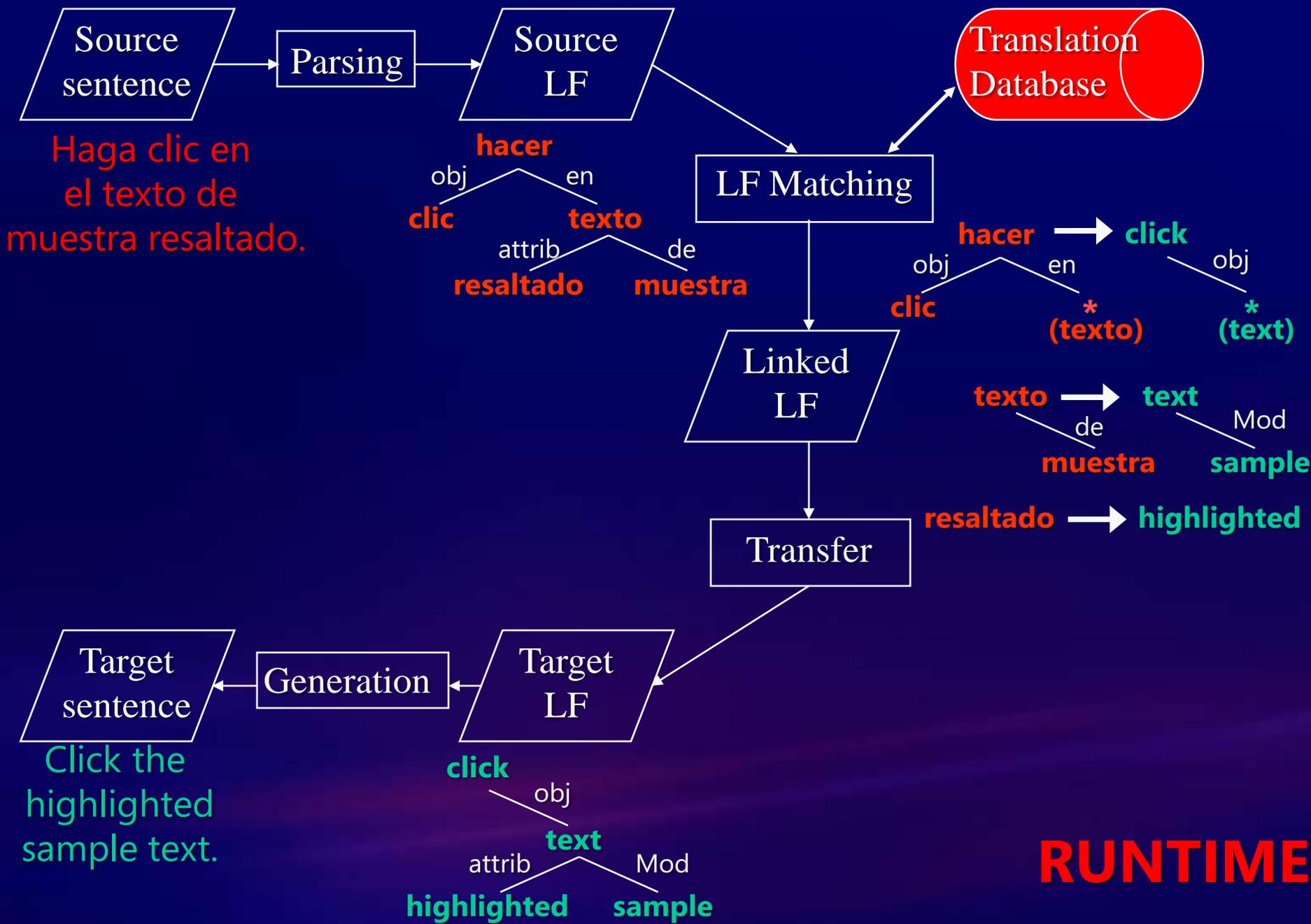
- A Little Background
- Evolution of Microsoft MT technology from rule-based to statistical hybrid
- Experiences with post-editing MT
- Experiences using raw MT
- Deployment of Windows Live Translator

A Little Background

- Natural Language Processing was the first group at Microsoft Research in 1991
- Initial efforts focused on rule-based parsing and grammar checking
 - Technology shipped in Word 97 grammar checker
- Emphasis on deeper “Logical Form” processing, Mindnet, & information retrieval applications followed
- Using NL components created in 6 languages, work on machine translation began in 1999
- First external visibility of MT was Customer Support KB pilot in 2002
- Today, ongoing NL research includes paraphrasing, sentiment analysis, ESL grammar & style checking, summarization, parsing, and of course, machine translation

Microsoft Research Machine Translation: MSR-MT, Version 1

- Began with a Logical Form-based EMBT system
 - Presented at EMBT workshop in 2001 and in Chapter 15 of "Recent advances in EMBT" (Carl & Way, ed.)
- Training
 - Parsing of parallel text in both languages, obtaining parallel logical forms (LFs)
 - Rule-based alignment of LFs
 - Extraction of paired, aligned LF fragments
- Runtime
 - Parse source language text, obtain LF
 - Match LF-fragments extracted during training
 - Choose best fragments based on size and frequency
 - Assemble a target LF
 - Generate target sentence from LF



Haga clic en el texto de muestra resaltado.

Click the highlighted sample text.

RUNTIME

MSR-MT, Version 1

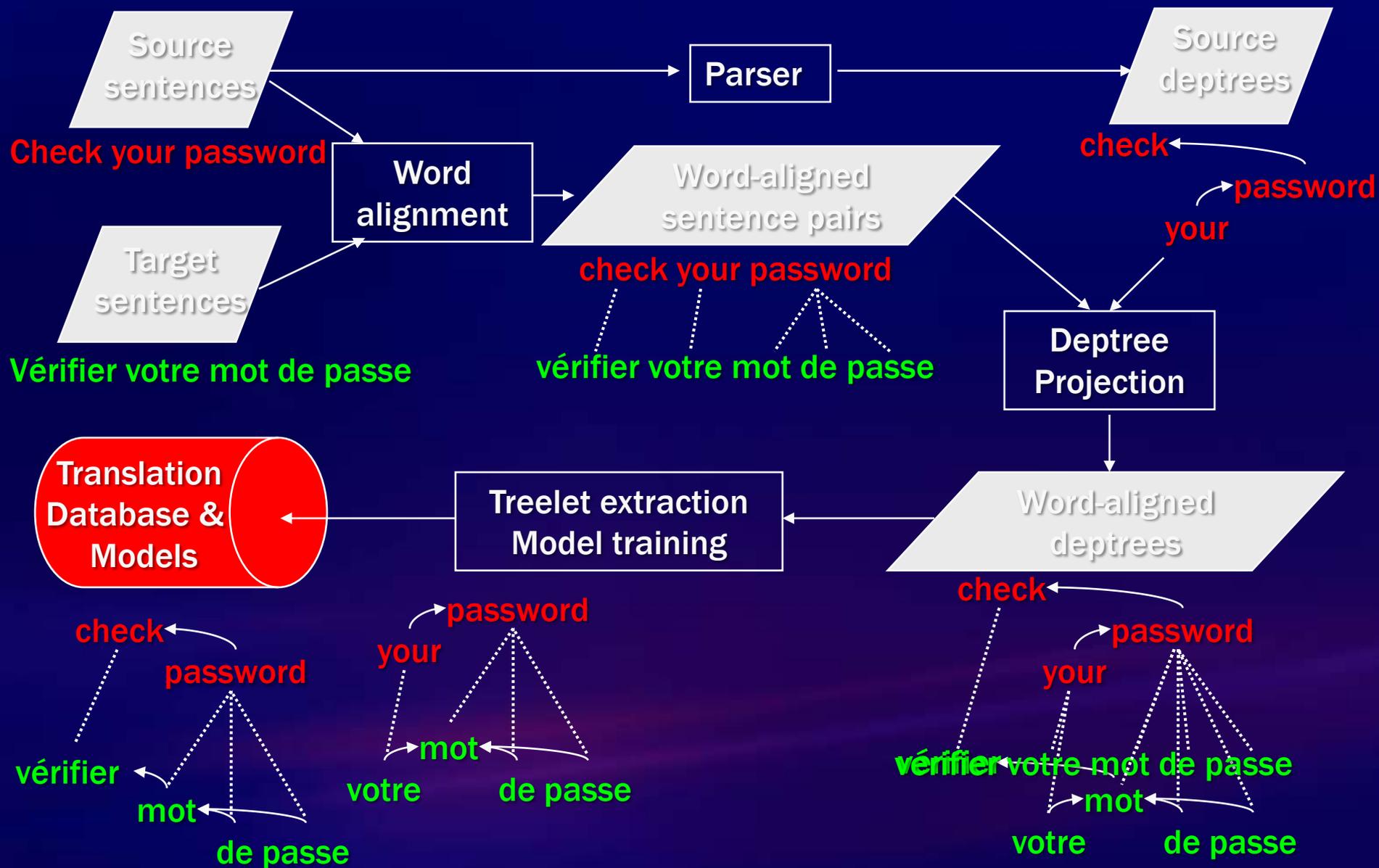
- Logical form representation:
 - Brought semantically related elements together
 - Handled superficially discontinuous but semantically connected 'phrases'
 - Allowed easy generalization of phrases, e.g. by inserting extra modifiers
 - Handled global re-ordering by simple movement/rotation in the LF structure
- BUT:
 - Difficulties dealing with combinatorial explosion of phrases
 - Heuristic scoring was brittle & hard to tune; no smoothing in the absence of examples
 - Generated text not as fluent without target language modeling

MSR-MT, Version 2:

Treelet translation

- Representative of current focus in MT research community
 - One big difference: emphasis on $E \rightarrow X$, rather than $X \rightarrow E$
- Combines advantages of linguistic processing with Phrasal SMT techniques
 - Exploits version 1 rule-based source language parser or newer dependency parsers
 - Uses rules for certain closed-class phenomena
 - Statistically aligns words
 - Employs an end-to-end statistical decoder
 - Computes a weighted combination of channel, ordering, and target language models
- Use source-language parser everywhere
 - Learn phrases in the tree
 - Dependency tree brings related words together
 - Handles discontinuous phrases
 - Learns mappings from “treelets” to target language strings, thus minimizing target language requirements
 - Reorder with simple tree rearrangements
- Menezes & Quirk 2005; Quirk et al., 2005

Training



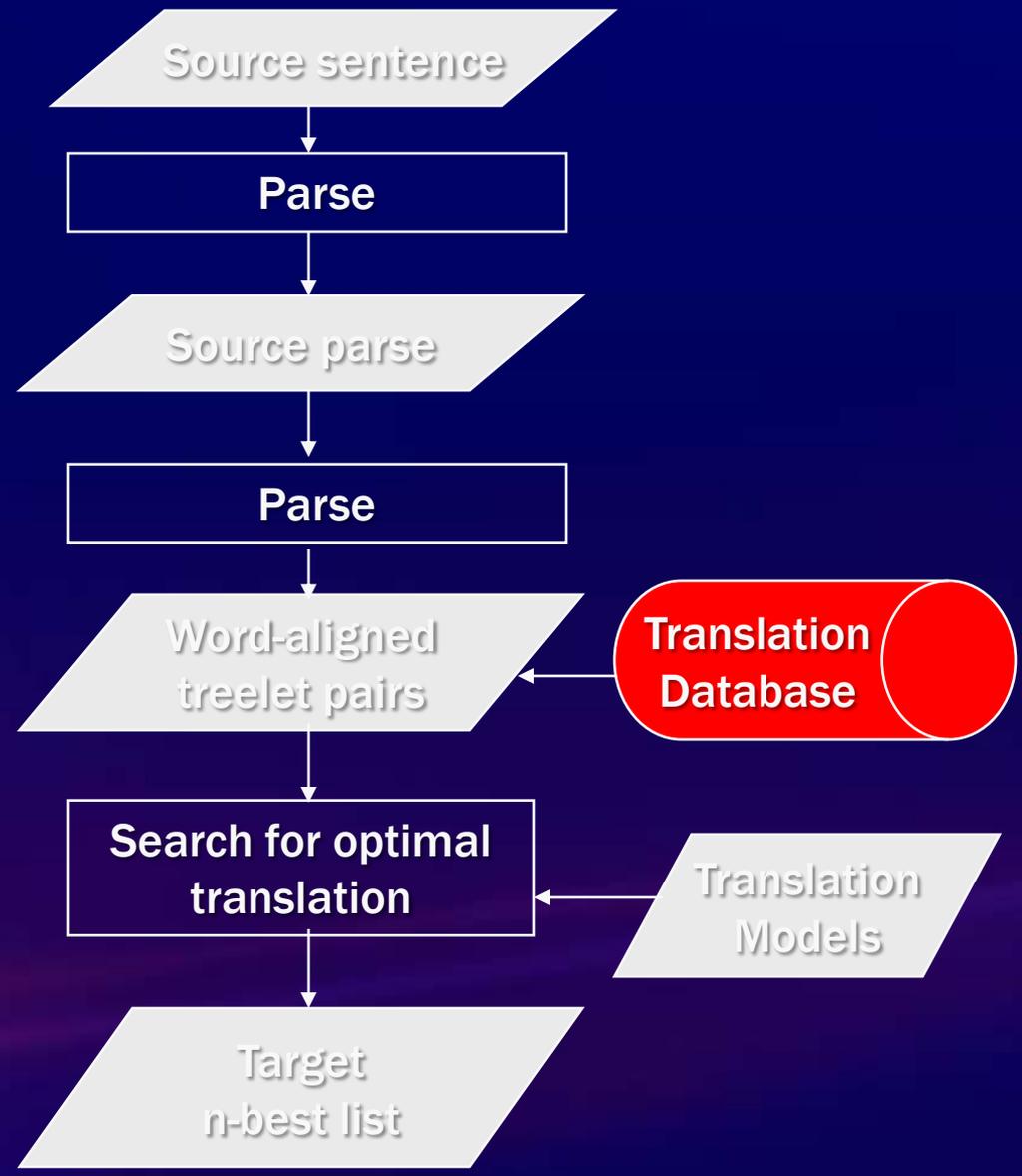
Runtime Decoding

Confirm the new password

confirm ←
password →
the new



confirmer ←
mot ←
le nouveau de passe



Ongoing research on Version 2

- Incorporate deeper syntactic information
 - So far, only part-of-speech and dependency structure
 - Current work on morphological modeling for richly inflected languages
 - Incorporate linguistic features into other models, not just ordering
- Exploring alternatives for supporting for $X \rightarrow E$ processing
- Performance and scaling
 - Recent significant speed ups in decoding
 - Larger language models

Microsoft Internal MT Service

- MSR-MT service runs a cluster of servers on our corporate intranet
- Trained with 5M to 50M words of previously translated Microsoft technical documentation
- From English to 25 languages, from 6 languages to English
- Extensive infrastructure to retrain & test languages every day or every few days on clusters of computer servers
- Available generally via the Word task pane & our internal web site
- Service API can be called from other applications as well

Currently available Languages

English ↔ French	English → B. Portuguese	English → Czech
English ↔ Spanish	English → Russian	English → Polish
English ↔ Japanese	English → Korean	English → Greek
English ↔ German	English → Dutch	English → Turkish
English ↔ Italian	English → T. Chinese	English → Hungarian
English ↔ S. Chinese	English → Arabic	English → Finnish
	English → Hebrew	English → Hindi
	English → E. Portuguese	English → Thai
	English → Swedish	
	English → Danish	
	English → Norwegian	

Using MT at Microsoft

Post-edited MT

- Users see human-edited MSR-MT translations
- Enables increased efficiency (and cost savings) in localization efforts
- Used when accuracy is critical
 - e.g., MT output is post-edited by localizers during translation of documentation and software strings

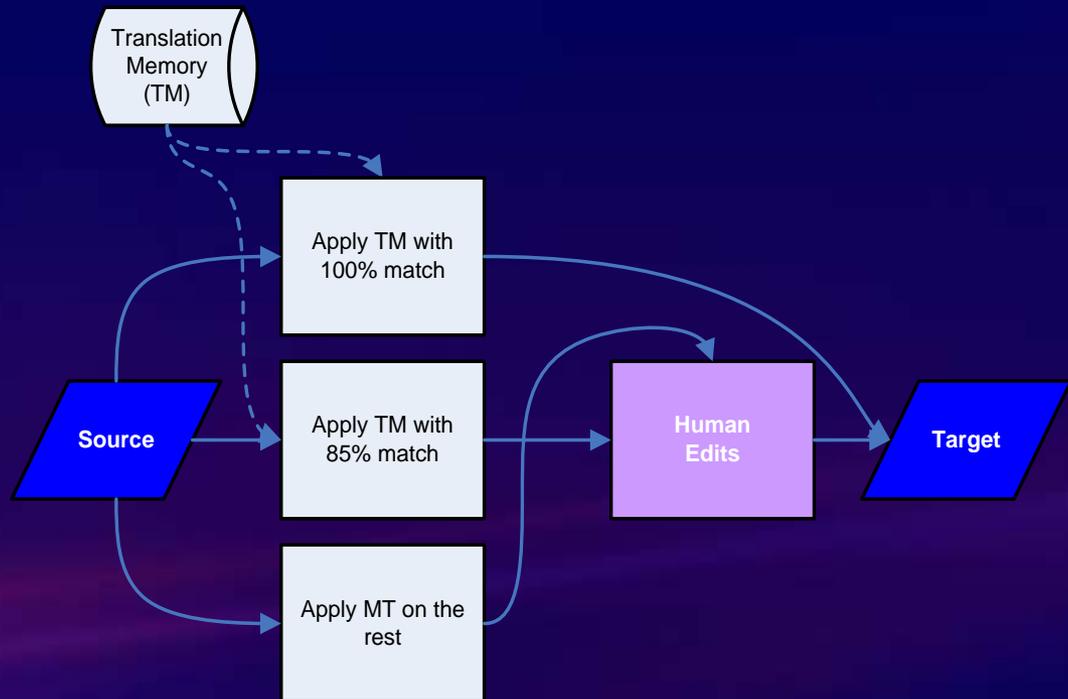
Raw MT

- Users see unedited MSR-MT translations
- Enables translation of material that otherwise goes untranslated
- Used when some errors can be tolerated
 - E.g., when the alternative is nothing at all: the CSS knowledge base
 - Users are motivated!

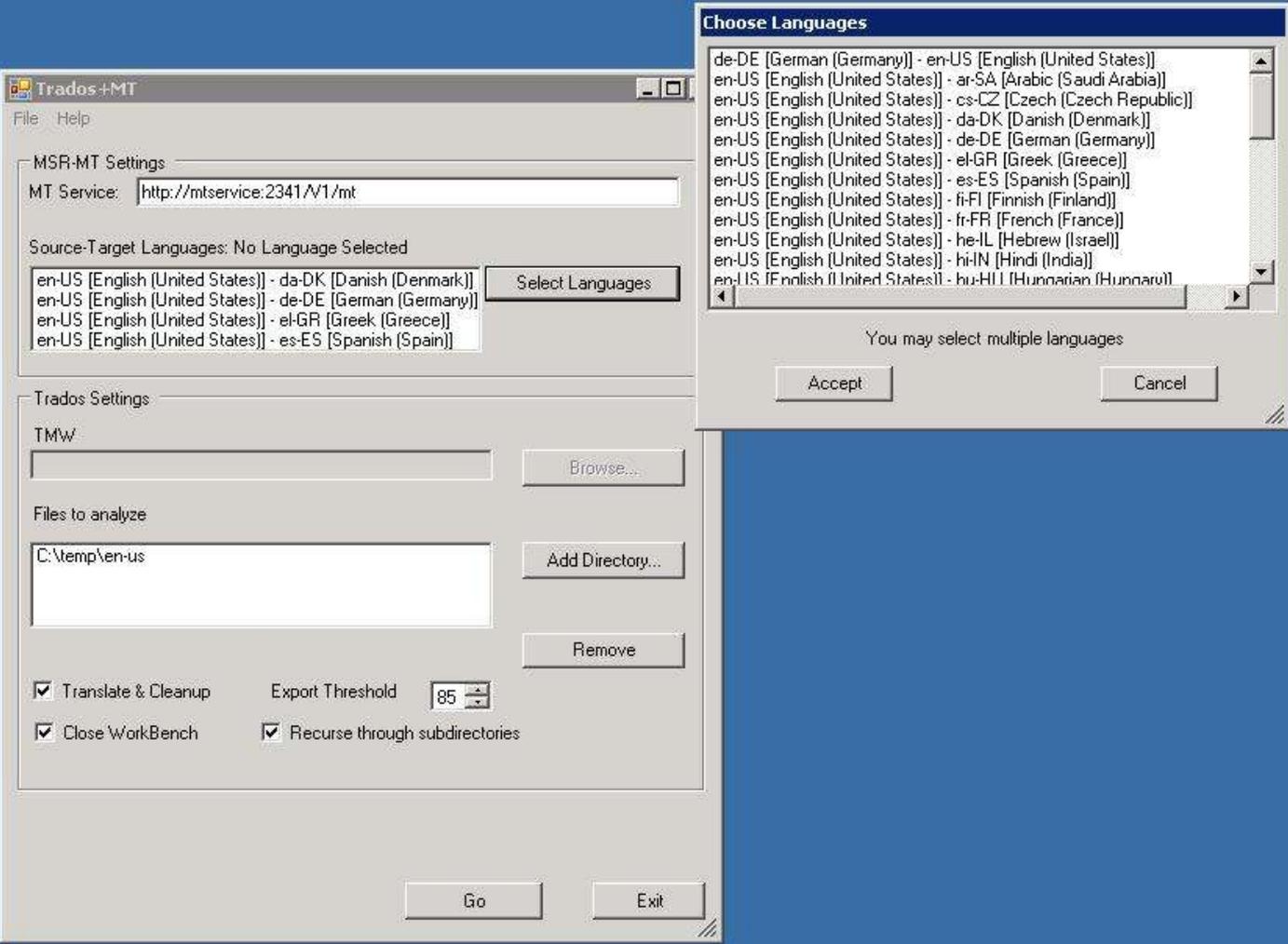
Microsoft Internal Use: Post-edited MT

- Many Microsoft Localization teams have piloted MT and continue to do so:

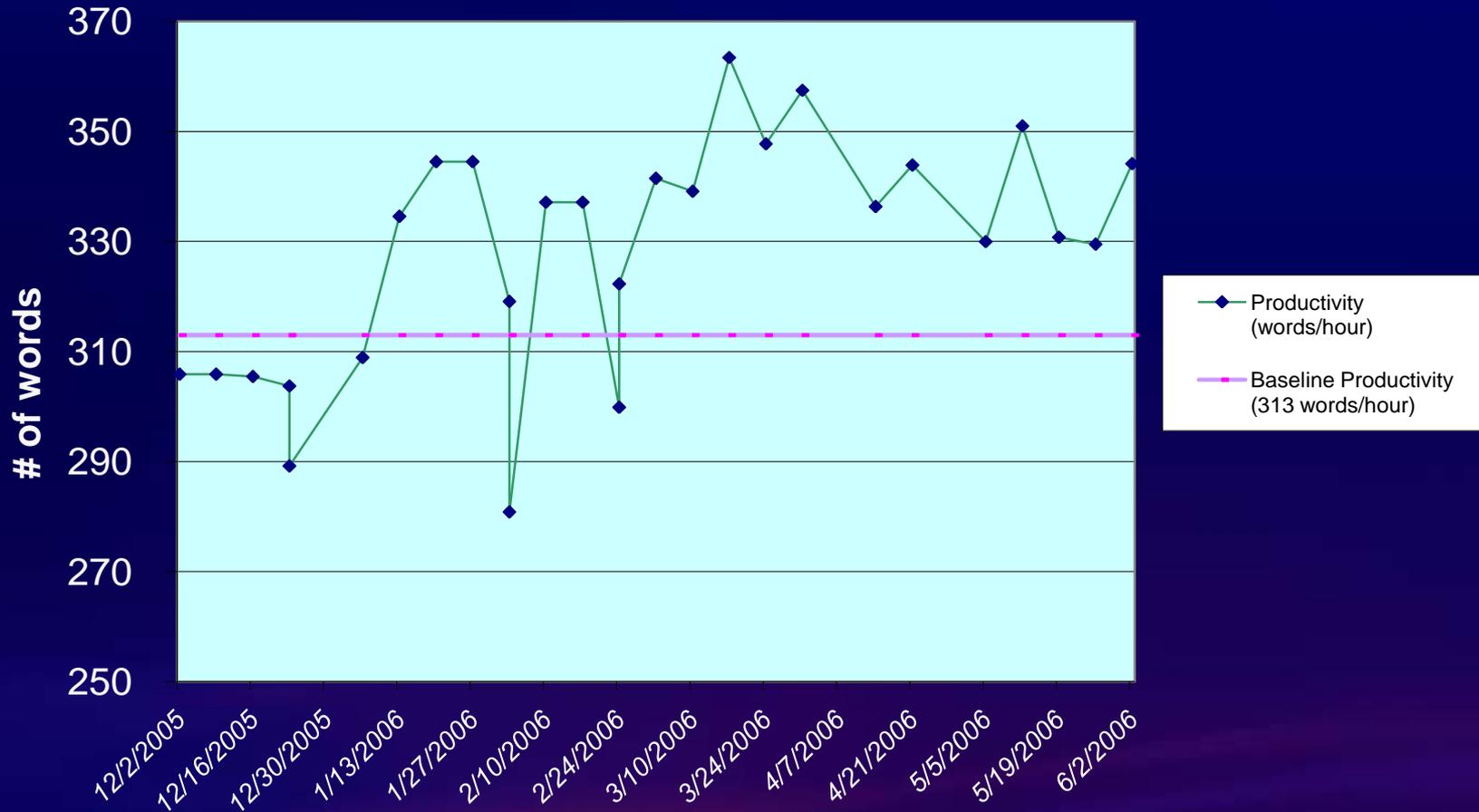
- Office 2007
- Windows Vista
- Visual Studio 2005
- Windows Server
- SQL Server
- Exchange
- Microsoft.com



SDL/Trados+MT: Integration of Microsoft MT with Translation Memory



MT Post-Editing Productivity



- Ranges from 5% to 25% savings in translation time/cost
- Depends on division of labor, post-editing quality guidelines, translator training, and vendor

Microsoft Internal Use: Raw MT

- Customer Support KB since 2002
- Last Windows Vista beta shipped with all help files machine translated into German and Japanese
- Visual Studio 2005 Brazilian Portuguese wiki allows user-developers to edit machine translated documentation and help
- Microsoft Developer Network (MSDN) is piloting machine translated articles in 7 languages

Customer Support Knowledge Base

- Knowledge Base (KB) Size
 - 200,000 articles x 600 words/article = approximately 120m words
 - 2,000+ articles change per week
- Previously, only 5%-20% of articles had been translated (with multi-million dollar annual budget)
 - Average cost to translate 1 article = \$150 (for European languages)
 - Complete, up-to-date translation not economically feasible (approx. \$30M initial outlay for one language, not including weekly updates)
- Customer satisfaction numbers for machine translated articles is comparable to & sometimes exceeds original English!

Number of articles

<u>Language</u>	<u>HT</u>	<u>MT</u>	<u>Total</u>
German	26k	176k	202k
Spanish	10k	185k	195k
French	33k	178k	211k
Japanese	64k	153k	217k

Customer Support Knowledge Base

PPT: Cómo determinar y/o cambiar el Estilo de viñeta - Microsoft Internet Explorer

Address: <http://support.microsoft.com/kb/132189/es>

España Microsoft.com/Spain Home | Mapa del sitio

Microsoft

Ayuda y Soporte Técnico

Página de inicio | Seleccione un producto | Buscar | Grupos de noticias | Contáctenos

PPT: Cómo determinar y/o cambiar el Estilo de viñeta

[Ver los productos a los que se aplica este artículo](#)

AVISO: Gracias por utilizar el servicio de Traducción Automática. Este artículo ha sido traducido por un sistema informático sin ayuda humana (Machine Translation). Microsoft ofrece estos artículos a los usuarios que no comprendan el inglés, exclusivamente, con el fin de que puedan entenderlos más fácilmente. Microsoft no se hace responsable de la calidad lingüística de las traducciones ni de la calidad técnica de los contenidos de los artículos así como tampoco de cualesquiera problemas, directos o indirectos, que pudieran surgir como consecuencia de su utilización por los lectores.

Selecciona idioma
Alemán

Id. de artículo : 132189
Última revisión : martes, 16 de agosto de 2005
Versión : 2.1

En esta página

- Resumen
- Más información
- Método 1: Seleccione algo a no ser [texto normal]
- Método 2: Cambie la fuente al principio de la línea

Resumen

La fuente del primer carácter de una línea con viñeta determina la fuente utilizada para la viñeta cuando utiliza viñetas de Texto de Normal. Este comportamiento es normal para la característica de viñeta en PowerPoint.

Si hace clic en Viñeta en el menú Formato y si hace clic en [Texto de Normal] en la lista Viñetas De, el diagrama de viñetas disponibles se dibuja utilizando la fuente del primer carácter del párrafo actualmente seleccionado.

Esto significa que si cambia la fuente que está utilizando en una lista con viñetas, también se cambia la viñeta.

[Volver al principio](#)

Más soporte técnico

- Contactar con Microsoft
- Presentaciones Técnicas
- Grupos de noticias

Herramientas

Imprimir la página

Internet

start

Inbox - Microsoft Out... My Documents

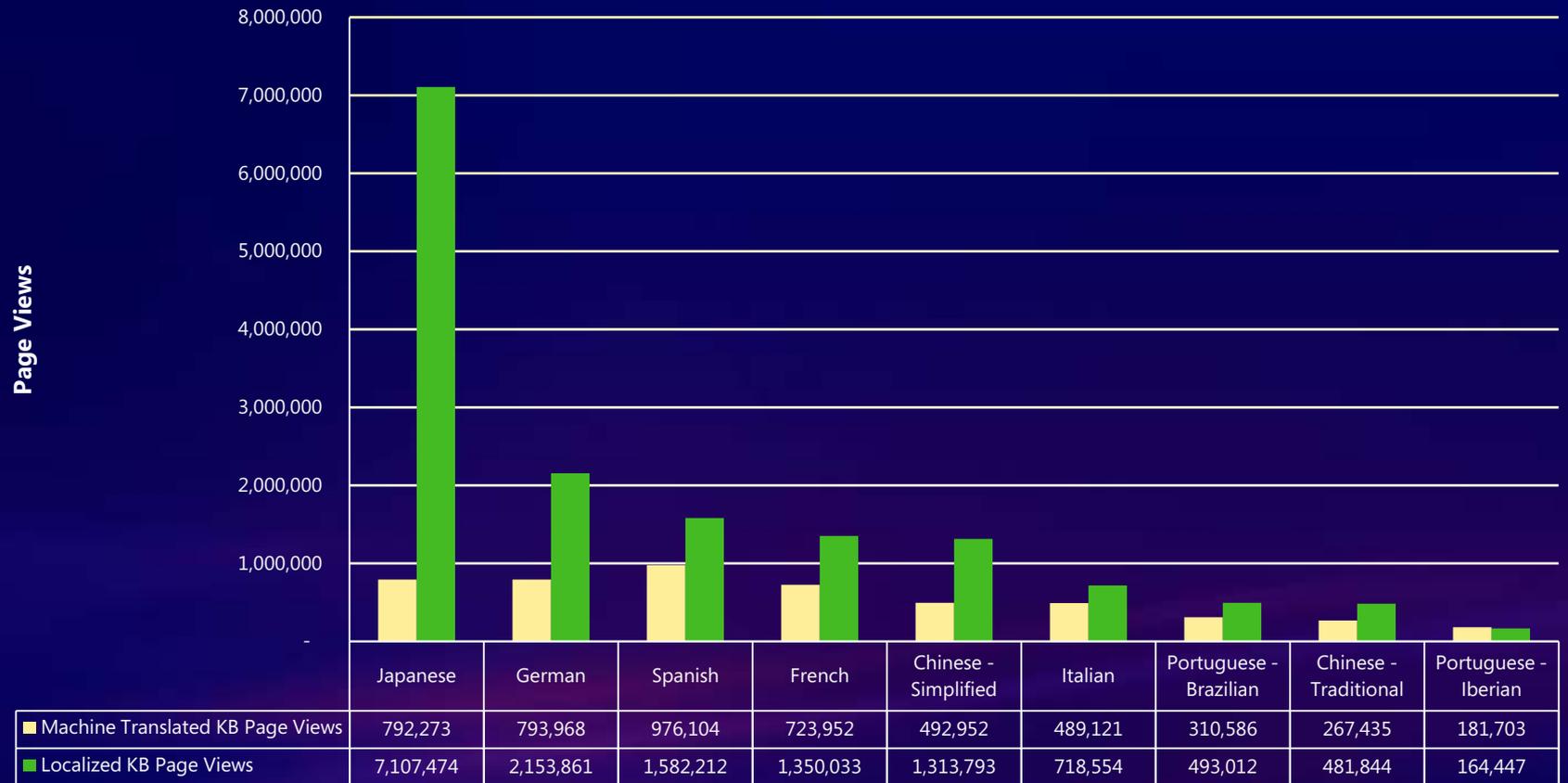
Machine Translation f... PPT: Cómo determina...

Search Desktop

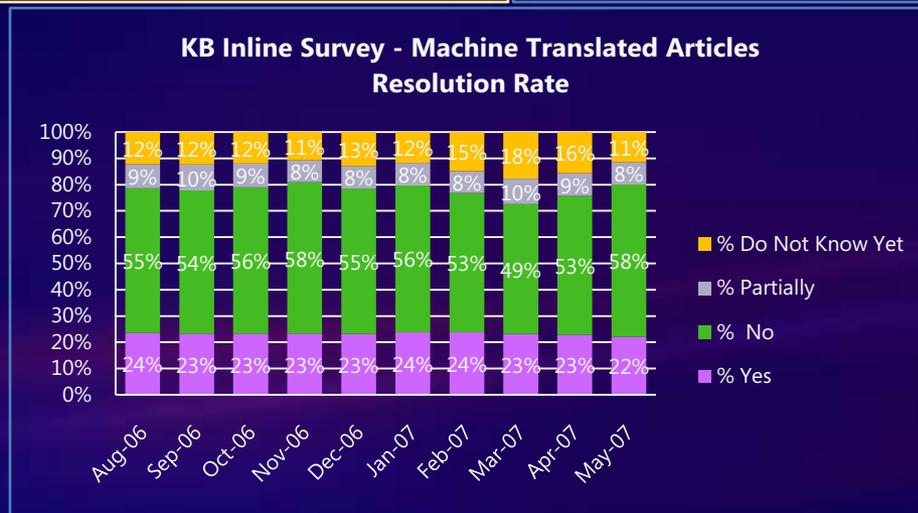
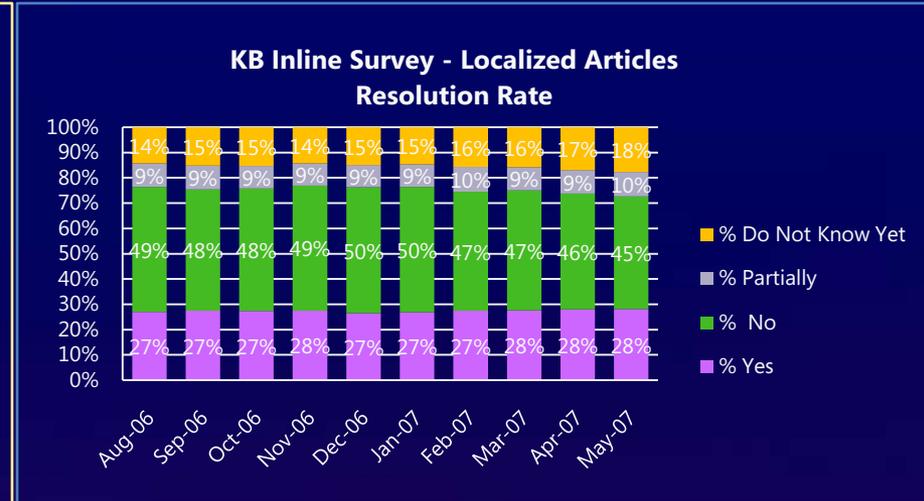
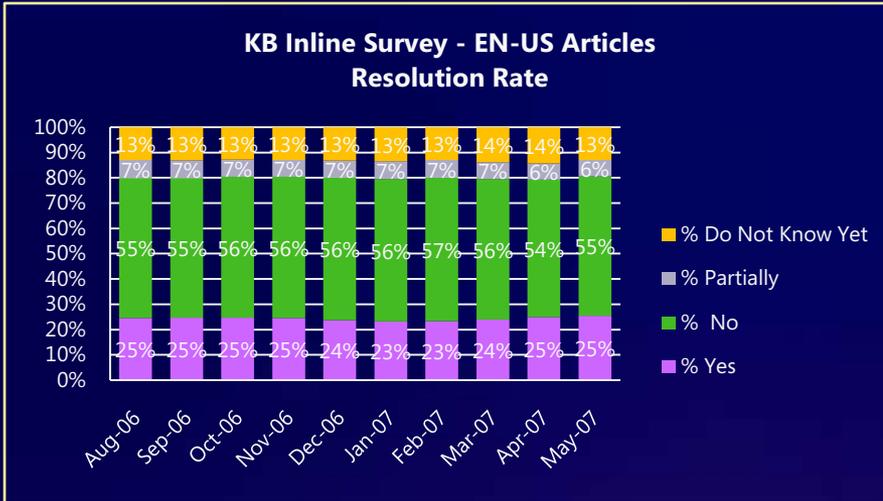
5:53 PM Thursday 5/11/2006

Customer Support KB: Page Views by Language

GSS KB Page Views for MT Languages - May 2007

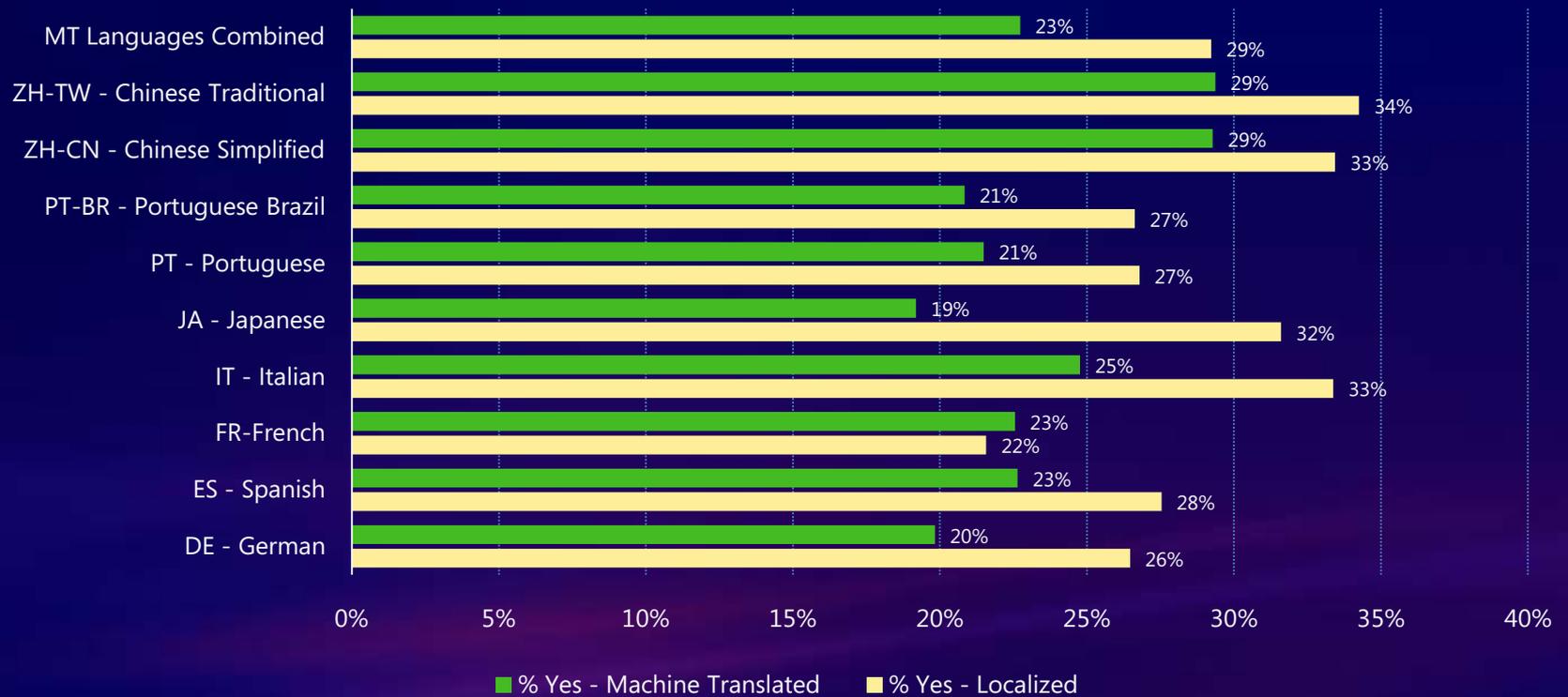


Customer Support KB: Resolution Rate over time for English, Human Translations, and Machine Translations



Customer Support KB: Resolution Rate by Language for Human Translations and Machine Translations

Survey 1361 - % Yes for Help Resolve Problem
Question For MT Languages
March to May 2007

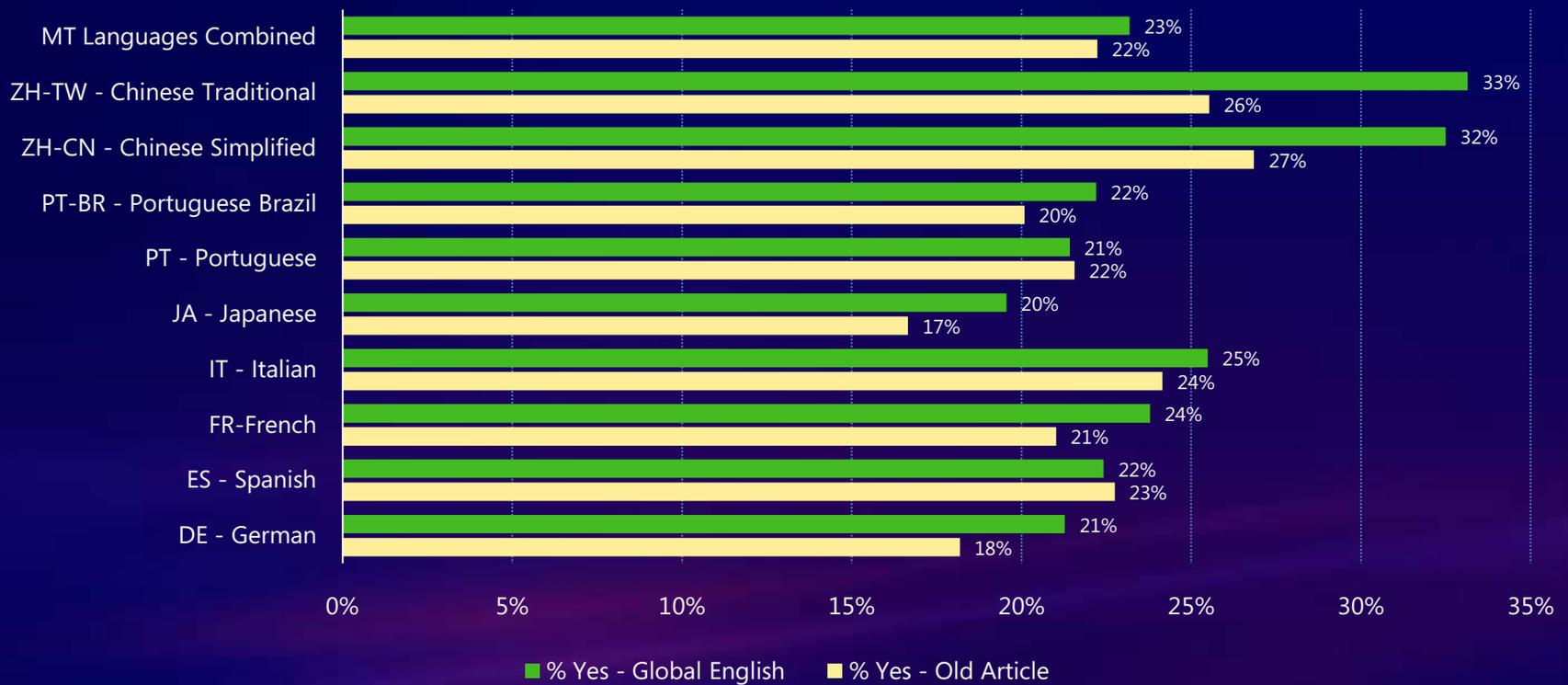


"Global English"

- Microsoft groups are using established authoring guidelines to improve MT quality (see Aikawa et al. 2007 – this afternoon)
- **EXAMPLE RULE: End every sentence at the first logical opportunity**
 - You can manually add root hints by using the DNS snap-in, by replacing the cache.dns file on your hard disk with the backup Cache.dns file, or by replacing it with the original version of the Cache.dns file from the Windows Server 2003 CD.
 - (MT) *Mit Hilfe, indem die cache.dns-Datei auf ihrer Festplatte mit der Sicherung-cache.dns-Datei ersetzt oder es durch die Originalversion der cache.dns-Datei von der Windows Server 2003-CD ersetzt, des DNS-Snap-in können Sie Stammhinweise manuell hinzufügen.*
- **You can easily turn this one long sentence into a bulleted list:**
 - You can manually add root hints in three ways:
 - You can use the DNS snap-in.
 - You can replace the Cache.dns file on your hard disk with the backup Cache.dns.
 - You can replace the Cache.dns file with the original version of the Cache.dns file from the Windows Server 2003 CD.
 - (MT) *Manuell können Sie Stammhinweise auf drei Arten hinzufügen:*
 - *Sie können das DNS-Snap-In verwenden.*
 - *Sie können die Cache.dns-Datei auf ihrer Festplatte durch die Sicherungskopie Cache.dns ersetzen.*
 - *Sie können die Cache.dns-Datei durch die zu erhaltende ursprüngliche Datei-Cache.dns-Version von der Windows Server 2003-CD ersetzen.*

Customer Support KB: Impact of Global English on Resolution Rate

Survey 1361 - % Yes for Help Resolve Problem
Question For MT Languages
March to May 2007

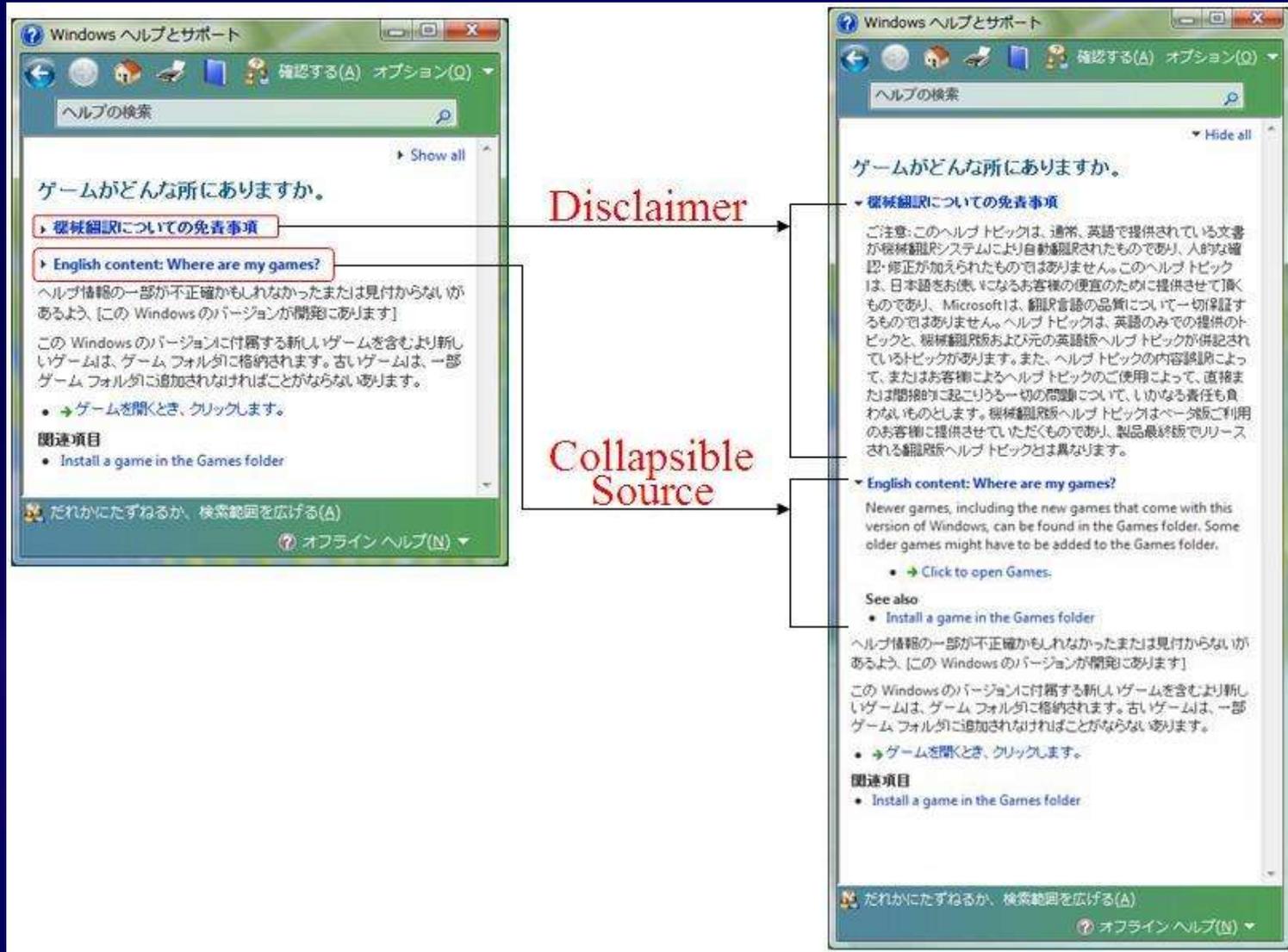


Customer Support KB: Number of Customers Helped by MT

ESTIMATED NUMBER OF CUSTOMERS HELPED BY MT: FY'07 TO DATE

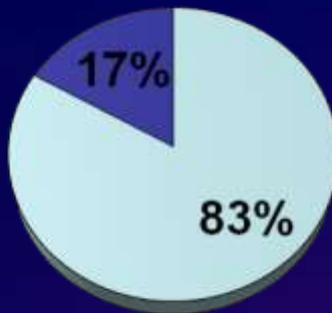
Language	MT KB Article	KB Resolution: % Yes for MT	Estimated # Customers Helped
	Page Views Jul 2006-May 2007		
DE: German	38,990,066	20%	7,798,013
FR: French	28,925,940	23%	6,652,966
IT: Italian	15,456,194	25%	3,864,049
JA: Japanese	84,198,014	19%	15,997,623
PT: Portuguese	2,577,645	21%	541,305
PT-BR: Portuguese-Brazil	8,091,415	21%	1,699,197
ZH-CN: Chinese Simplified	22,506,847	29%	6,526,986
ZH-TW: Chinese Traditional	9,196,373	29%	2,666,948
TOTAL			45,747,087

Windows Vista Beta



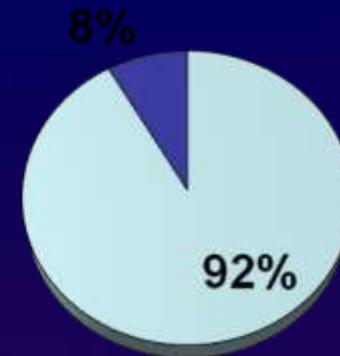
Japanese MVP Customers consider MT valuable

Did MT Help?



■ Yes ■ No

Do you want more machine translated content?



■ Yes ■ No

MSDN: Visual Studio 2005 Brazilian Wiki

<http://msdnwiki.microsoft.com/pt-br/mtpswiki/default.aspx>

MSDN Wiki Home Page - Windows Internet Explorer

http://msdnwiki.microsoft.com/pt-br/mtpswiki/default.aspx

File Edit View Favorites Tools Help

3rd Party Component Reque... MSN Hotmail - Inbox Location MSDN Wiki Home Page

Microsoft.com Brasil Home | MSDN Wiki Home | Comentários | fa

Procurar na documentação:

MSDN Wiki

MSDN Library

- Visual Studio
- .NET Framework SDK

Bem-vindo à versão Beta da MSDN Wiki em Português

Este site contém dois tipos de funcionalidade e conteúdo:

- Uma seção com toda a documentação do Visual Studio 2005 e do .NET Framework 2.0 traduzida automaticamente. Você pode editar esta seção para sugerir alterações que aprimorem a tradução. Moderadores irão revisar suas sugestões e as adicionar à tradução oficial.
- A seção com o conteúdo da comunidade permite que você adicione exemplos de programas, dicas e outras informações interessantes.

Para esclarecer suas dúvidas ou conhecer mais esse site, visite a seção de perguntas e respostas. Para reportar problemas, dar sugestões ou saber mais sobre nossos planos futuros, visite a área de trabalho da MSDN Wiki.

Selecione um item da biblioteca à esquerda para começar.

Tópicos editados

- [Assistente de controle ActiveX MFC](#)
Alterado por Ricardo Oneda -- Conceitos do Visual C++: criando e gerenciando projetos Assistente de controle ActiveX...
- [Representantes no Visual Basic](#)
Alterado por Ricardo Oneda -- Conceitos da linguagem Visual Basic Representantes no Visual Basic ...
- [Ferramentas do Visual Studio para o Office](#)
Alterado por Ricardo Oneda -- Visual Studio Tools for the Microsoft Office System: Ferramentas do Visual Studio para...
- [Permissões de acesso a código](#)
Alterado por Ricardo Oneda -- .NET Framework Developer's Guide Permissões de acesso a código Perm...
- [Segurança relacionados de instrução](#)

Estatísticas da comunidade

Blocos de conteúdo: 163
Alterações de conteúdo: 213
Tópicos com conteúdo da comunidade: 160
Contribuidores: 30

Contribuidores mais ativos

- [Ricardo Guerra \(85\)](#)
- [Ricardo Oneda \(29\)](#)
- [roberto_sonnino \(21\)](#)

Nos últimos sete dias

- [Ricardo Guerra \(1\)](#)
- [dasarobr \(1\)](#)
- [Alexandro Prado \(1\)](#)

Blog da MSDN Wiki

- [MSDNWiki: Recent Edits RSS](#)
- [MSDNWiki: Community Content RSS launch](#)
- [MSDN Wiki migration to MSDN2](#)
- [MSDN Wiki Project Update](#)
- [Enhanced RSS Feed for MSDN Wiki](#)

MSDN Library: Machine Translation Pilot

- 7 languages: B. Portuguese, French, Italian, Japanese, Spanish, S. Chinese, T. Chinese
- Translated library articles displayed side-by-side with original text and synchronized highlighting of paragraphs

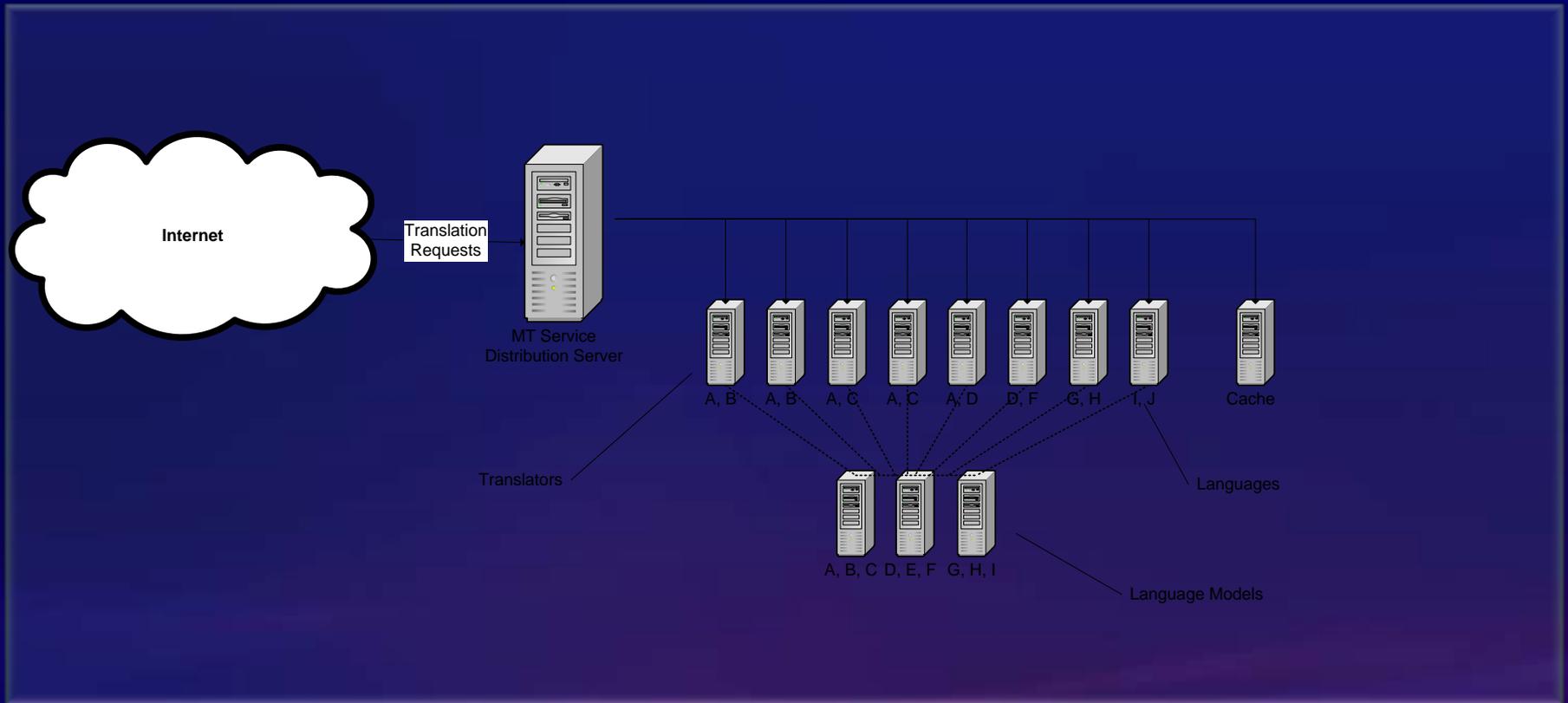
<http://mtbeta.msdn.microsoft.com/>



Windows Live Translator

- Now at: <http://translate.live.com>
- Internet translations in 13 bi-directional language pairs
- “Free” like other services: drives ad-revenue and increased traffic to related web sites
- Uses Microsoft Research’s own machine translation software for computer-technical text and web pages in 8 language pairs
- Currently uses Systran as default for all other domains and languages

Windows Live Translator Web Service Architecture



- Translation processing is distributed across multiple machines to achieve performance, scaling, & reliability

Windows Live Translator Bilingual Viewer user interface

- Focus on enhanced user experience
 - Mitigate MT shortcomings
 - Assist in language learning/understanding
- Four viewing modes:
 - Side-by-side
 - Top-bottom
 - Original text with hover translation
 - Translated text with hover original
- Synchronized scrolling and highlighting
- Progressive rendering of translated text
- Bilingual navigation of hyperlinks

<http://translate.live.com>

Windows Live Translator Bilingual Viewer user interface

The screenshot displays the Windows Live Translator Bilingual Viewer interface within a Windows Internet Explorer browser window. The browser's address bar shows the URL: <http://www.windowstranslator.com/BV.aspx?MCT=en-US#http://msdn2.microsoft.com/en-us/vbasic/bb736015.aspx>. The interface is split into two main columns: 'Original: English' on the left and 'Translated: Spanish' on the right. The original text is from the MSDN 'Visual Basic Developer Center' and is titled 'Getting Started with Windows Communication Foundation'. The translated text is in Spanish, titled 'Introducción a Windows Communication Foundation'. Below the text, there are two side-by-side 'New Project' dialog boxes, each showing a list of project types (Visual Basic, Visual C#, Visual J#, Visual C++) and templates (Visual Studio installed templates, Blank Solution, My Templates). The footer of the browser window includes copyright information for Microsoft (© 2007), links for Privacy and Legal, and a note that the translation is powered by Microsoft Research.

Translator Beta - Windows Internet Explorer
http://www.windowstranslator.com/BV.aspx?MCT=en-US#http://msdn2.microsoft.com/en-us/vbasic/bb736015.aspx

Windows Live™
Translate URL: <http://msdn2.microsoft.com/en-us/vbasic/bb736015.aspx> English - Spanish

Original: English [Go to original page](#) **Translated: Spanish** [Add to Favorites](#) [Is this translation helpful?](#)

Welcome | Sign In | Add LIVE SEARCH to your Browser! | United States - English | Microsoft.com
msdn Visual Basic Developer Center
Search MSDN with Live Search Home Library Learn Downloads
Blogs Webcasts Related Sites Featured MVPs Visual Basic Team
Printer Friendly Version Add To Favorites Send
Click to Rate and Give Feedback
MSDN > Developer Centers > Visual Basic Developer Center > Community

Getting Started with Windows Communication Foundation

By [Rob Windsor](#), Visual Basic MVP, ObjectSharp Consulting

Windows Communication Foundation (WCF) is Microsoft's next-generation platform for building distributed systems. It was released as part of the .NET Framework 3.0 and is designed to consolidate and extend the APIs from previous versions of the Framework (i.e., ASP.NET Web Services, .NET Remoting, Enterprise Services (COM+), and message queuing).

Getting started with WCF isn't difficult, but there are a couple steps along the way that may not be obvious. This article will walk you through the process of creating and consuming two simple services. To achieve this goal, we need to discuss service classes, service hosts, service configuration, and client proxies.

To follow along with the walkthrough, you'll need [Visual Studio 2005](#), the [.NET Framework 3.0](#), and the [Visual Studio 2005 extensions for .NET Framework 3.0 \(WCF and WPF\)](#) installed.

To start, create an empty solution called **GettingStartedWithWCF**.

New Project

Project types:	Templates:
<ul style="list-style-type: none">Visual BasicVisual C#Visual J#Visual C++Distributed System Solutions	Visual Studio installed templates <ul style="list-style-type: none">Blank SolutionMy Templates

Bienvenido | [Iniciar sesión en](#) | [Agregar VIVO Search a su explorador.](#) | Estados Unidos - Inglés | Microsoft.com
msdn Centro para desarrolladores de Visual Basic
Search MSDN with Live Search Principal Biblioteca Obtenga informacón
Blogs Webcasts Sitios relacionados MVP destacado Equipo de Visual Basic
Versión descriptiva de la impresora Agregar A favoritos Enviar
Haga clic aquí para tipo y de a
MSDN > Desarrollador centros > Centro para desarrolladores de Visual Basic > La comunidad

Introducción a Windows Communication Foundation

Por [Rob Windsor](#), Visual Basic MVP, ObjectSharp Consulting

Windows Communication Foundation (WCF) es de Microsoft de próxima generación Plataforma para la creación de sistemas distribuidos. Se publicó como parte de .NET Framework 3.0 y está diseñado para consolidar y extender las API de versiones anteriores de Framework (es decir, los servicios Web ASP.NET, .NET Remoting, Enterprise Services (COM+) y la cola de mensajes).

Introducción a WCF no es difícil, pero hay un par de pasos a lo largo del modo que puede que no sea evidente. En este artículo le guiará a través del proceso de creación y el uso dos servicios simples. Para lograr este objetivo, necesitamos tratar las clases de servicio, servicio hosts, configuración del servicio y proxies cliente.

Para seguir junto con el tutorial, necesitará [Visual Studio 2005](#), el [.NET Framework 3.0](#) y las [extensiones de Visual Studio 2005 para .NET Framework 3.0 \(WCF y WPF\)](#) instaladas.

Para iniciar, crear una solución vacía llama **GettingStartedWithWCF**.

New Project

Project types:	Templates:
<ul style="list-style-type: none">Visual BasicVisual C#Visual J#Visual C++Distributed System Solutions	Visual Studio installed templates <ul style="list-style-type: none">Blank SolutionMy Templates

© 2007 Microsoft | Privacy | Legal Help | Account | Feedback | Translation powered by Microsoft Research

Microsoft Machine Translation: Lessons learned

- Real users are invaluable to real research
- User experience is as important as translation quality
- Seamless integration into existing software applications and user scenarios is key to broad acceptance
- We can never have enough clean, parallel data
- Having lots of parallel, distributed computing power makes a difference
- Really good machine translation is still 5 years away!

Thank you