

## **Wired for Peace and Multi-language Communication: Virtual Diplomacy in Northeast Asia and Web-based Translation**

Vladimir Oboronko  
*Wired for Peace* Project Coordinator  
University of California  
Institute on Global Conflict and Cooperation  
9500 Gilman Drive  
(Robinson Complex #1229, cor. Scholars Dr. N. & Int'l Lane)  
La Jolla, CA 92093-0518  
phone (858) 534-0819  
fax (858) 534-7655

**IGCC** <http://vwww-igcc.ucsd.edu/>

**W4P** <Http://www-neacd.ucsd.edu/>

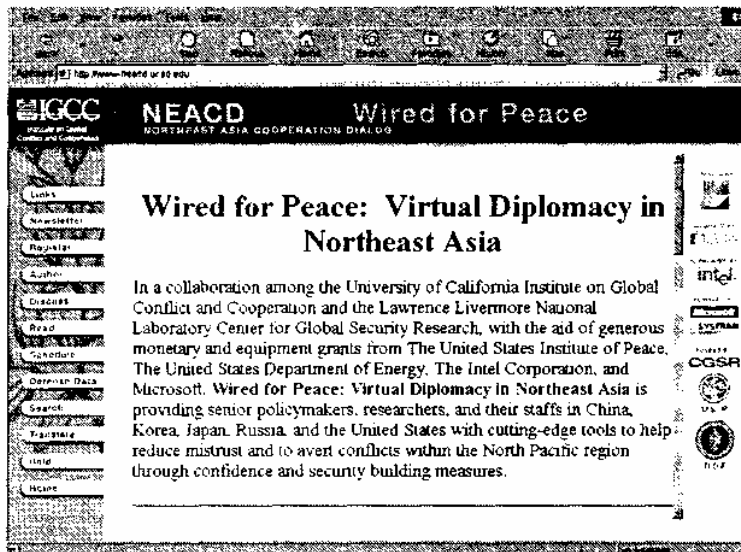
**email** [voboronk@uc.sd.edu](mailto:voboronk@uc.sd.edu)

### *Abstract*

Our project *Wired for Peace: Virtual Diplomacy in Northeast Asia* (<Http://www-neacd.ucsd.edu/>) has as its main aim to provide policymakers and researchers of the U.S., China, Russia, Japan, and Korea with Internet based tools to allow for continuous communication on issues of the regional security and cooperation. Since the very beginning of the project, we have understood that Web-based translation between English and Asian languages would be one of the most necessary tools for successful development of the project. With this understanding, we have partnered with Systran ([www.systransoft.com](http://www.systransoft.com)), one of the leaders in MT field, in order to develop Internet-based tools for both synchronous and asynchronous translation of texts and discussions. This submission is a report on a work in progress.

### *1. Background*

*Wired for Peace* is an international project on virtual diplomacy. Its goal is to organise regular communication between policymakers and researchers of Northeast Asia. For this purpose, we have created a W4P website (<http://www-neacd.ucsd.edu/>) that offers Internet-based communication and information-sharing tools for the participants of the program.



*Wired for Peace* (W4P) is a project functioning within a bigger project of Northeast Asia Cooperation Dialogue (NEACD). Under the auspices of NEACD, policymakers and researchers from the U.S., China, Russia, Japan, and Korea meet once a year and discuss issues of international security of the region. The W4P has been designed as offering online, in-between-the-meetings, communication among the NEACD members.

W4P has two major goals. One, a medium-term one, is to provide Web-based communication tools for NEACD members. The other, a longer-term one, is to involve more people in policymaking and academic communities of participating countries.<sup>1</sup>

Since the official language of NEACD is English, right now, basically almost every member of NEACD has English good enough to "hear and be heard." Meanwhile, for information-sharing and multi-lateral discussion purposes, we need to have an ability to translate documents to and from b/w English and every other language of NEACD participants: i.e. Chinese, Russian, Korean and Japanese.

In the longer run, if we want discussion and information-sharing regarding issues of security in the region to widen and include more and more members of Northeast Asia policymaking and research communities, we need to be able to use relatively simple and effective tools of machine translation. We have already felt a strong need for such tools; even now, in the beginning stages of our W4P project development.

Thus, our China Country Manager has been recently conducting W4P training in Beijing of our Chinese colleagues. She was told that one of the factors slowing down the Chinese project participants was that the experts on Northeast Asia felt that they were not proficient enough in English to participate in W4P site based discussions and communication. This finding simply confirms our initial understanding that to get ahead

<sup>1</sup>A widely known Japanese GLOCOM (Global Communication) project, <http://www.glocom.ac.jp/top>, is oriented towards as wide as possible an intellectual community and the general public on the whole. GLOCOM's focus is on issues regarding the future of information society; Japanese society is another main topic of GLOCOM. W4P is different from GLOCOM on both counts: 1) it involves efforts of policymakers and experts from North East Asian countries; as a rule, it is closed from the general public; and, (2) it focuses on the area of international security and relations in the region.

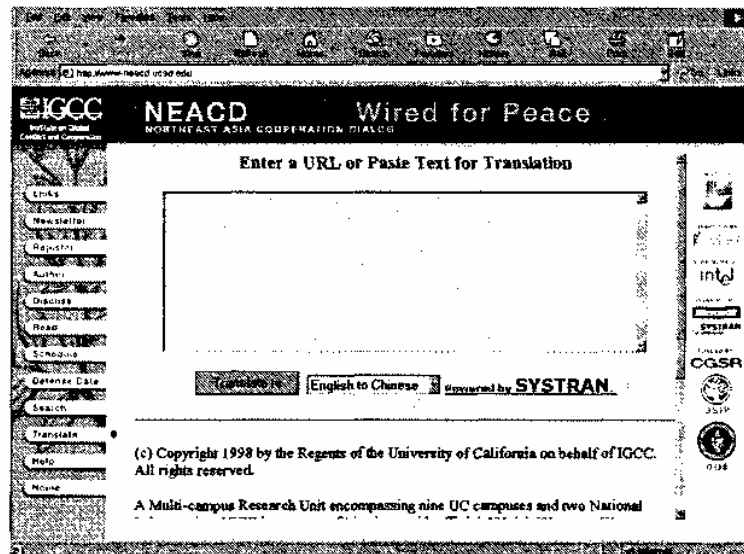
with the development of our international project *Wired for Peace*, we do need to provide our participants with simple enough and effective tools of Web-based translation.

With this goal in view, in the course of development and implementation of W4P project, we have partnered with Systran (<http://www.svstransoft.com/>), one of the foremost authorities in MT (including Web-based) translation.

## 2. Project Design and Implementation

For the sake of efficiency of our site and its tools, we assume that our "common lowest denominator" user has only very basic computer and Internet skills. Under this assumption, we need to integrate at our website a very user-friendly interface of multi-language translation. There are two major features that Systran has provided us with:

1. One feature is an option of seeing the whole W4P site in a language of your choice (from the set of NEACD participants' languages). The translation filters are set up in such a way that any English-language document that has been posted at our site gets translated into a language you've chosen upon entering the site.
2. The other feature is translation of a text into or from English through copying and pasting, or typing the text's URL in the translation window. In this case, the text is translated on-the-fly through a Systran provided link that is hooked up to a translation engine installed and maintained at the Systran's server.



This way, we have a translation interface simple enough to be used even by a person with a relatively low prior exposure to Internet.

Since we are looking at translation as a tool to enhance communication between members of the international project, we do not expect the Web-based translation to be perfect. Together with Systran, we are co-developing the translating tools that will allow

our members to participate in online discussions (both synchronous and asynchronous) and get acknowledged with relevant texts in NEACD participants' languages.

### 3. *Tuning the MT system on W4P texts*

The texts relevant to NEACD activity mainly belong to the field of international relations and security. The Systran's general purpose machine translation systems have not been tuned to translate this kind of texts. To improve the translation quality, we have assigned country managers to work with Systran developers.

Our country managers work on building Customer Specific Dictionaries in the field of international relations, security etc. Besides providing such domain specific words and expressions, we also provide feedback on translation performance. We have a set of procedures established b/w IGCC and Systran that allow us to improve quality of translation of texts through iterations of translations.<sup>2</sup>

---

<sup>2</sup> We are also working on calculating frequency distribution for words and expressions related to the field of international relations. We have come up with initial measure of effectiveness of translation and expect this measure to allow us to trace our progress in improving translation of texts related to the field of Northeast Asia security and cooperation. Please see the current statistics of improvement of translation in Appendix 1.

Article	Language	Duration	# of					SS/TSOA (%)	Exp/1,000 Words	Wds/1,000 Words	L1/1,000 Words
			SS <sup>1</sup>	Expression	Word	LF <sup>2</sup>	TSOA <sup>3</sup>				
pp35	Chinese	7/11-8/4	79	119	78	54	696	11.30	7.41	5.13	
	Japanese	7/11-8/4	106	112	79	46	696	10.64	7.50	4.37	
pp36	Chinese	8/11-8/17	187	175	76	170	1537	7.61	3.31	7.40	
	Japanese	8/12-8/21	241	245	116	487	1537	10.66	5.05	21.19	
pp37	Chinese	8/28-9/4	148	130	45	430	1392	3.59	1.24	11.89	
	Japanese	8/29-9/8	151	213	68	150	1392	5.89	1.88	4.15	
pp38	Chinese	9/8-9/15	47	108	20	190	1016	3.99	0.74	7.02	
	Japanese	9/11-9/18	77	153	35	91	1016	5.65	1.29	3.36	
pp39	Chinese	9/18-9/29	45	52	12	77	793	3.98	0.92	5.89	
	Japanese	9/23-10/						0.00	0.00	0.00	

- 1.SS: Senseless Sentences  
2.LF: Linguistic Feedback  
3.TSOA: Total sentences of the article