



**KantanMT.com**  
No Hardware. No Software. No Hassle MT.

**Lexically Constrained Decoding for Sequence Generation**  
Approaches to support Terms, Tags & Placeholders

# The Problem....

<b>original</b>	Hello <g id="1" ctype="x-bold;">World!</g>
<b>after tokenization</b>	Hello < g id = " 1 " ctype = " x-bold ; " > World ! < / g >
<b>after escaping</b>	Hello &lt; g id = &quot; 1 &quot; ; ctype = &quot; x-bold ; &quot; &gt; World ! &lt; / g &gt;
<b>after lowercasing</b>	hello &lt; g id = &quot; 1 &quot; ; ctype = &quot; x-bold ; &quot; &gt; world ! &lt; / g &gt;
<b>after translation</b>	guten tag &lt; code &amp;lt; ept id = &amp;quot; 1 &amp;quot; ; ctype = &quot; &quot; &quot; x-bold ; &quot; &gt; world money &lt; / g &gt; !
<b>after recasing</b>	Guten Tag &lt; Code &amp;lt; ept id = &amp;quot; 1 &amp;quot; ; ctype = &quot; &quot; &quot; x-bold ; &quot; &gt; World Money &lt; / G &gt; !
<b>after detokenization</b>	Guten Tag &amp;lt; Code &amp;lt; ept id = &amp;quot; 1 &amp;quot; ; ctype = "" "x-bold;" &gt; World Money &amp;lt; / G &gt; !
<b>after unescaping</b>	Guten Tag < Code &lt; ept id = &quot; 1 &quot; ; ctype = "" "x-bold;" > World Money < / G > !

# The Results...

ENGLISH - DETECTED   ENGLISH   SPANISH   FRENCH   ↕   POLISH   ITALIAN   ENGLISH

<tag1><tag2>Hello <strong style="type1" color="#873"> World</strong></tag2></tag1> ×   <tag1> <tag2> Ciao <strong style="type1" color="#873"> Mondo </strong></tag2></tag1> ✓

English   French

<tag1><tag2>Hello <strong style="type1" color="#873"> World</strong></tag2></tag1>   <tag1><tag2>Hello <strong style="type1" color="#873"> Monde</strong></tag2></tag1> ✗

English (USA)   Translate   Italian (Italy)

<tag1><tag2>Hello <strong style="type1" color="#873"> World</strong></tag2></tag1>   <tag1><tag2>Ciao<strong style="type1" color="#873"> Mondo</strong></tag2></tag1> ✓

<tag1><tag2>Hello <strong style="type1" color="#873"> World</strong></tag2></tag1> ×   <tag1><tag2>Hallo <strong style="type1" color="#873"> Welt</strong></tag2></tag1> ✗

# Strategy #1: Structure Forming

- Subtitling DFXP Format

```
<p begin="00:01:14:01" end="00:01:17:05">And if that is not enough world championship action,</p>  
<p begin="00:01:17:07" end="00:01:21:17">you can pick up the World Championship Pikachu Poké Plush</p>  
<p begin="00:01:21:19" end="00:01:23:02">at pokémoncenter.com.</p>
```

# Strategy #1: Structure Forming

- **Subtitling DFXP Format**

```
<p begin="00:01:14:01" end="00:01:17:05">And if that is not enough world championship action,</p>  
<p begin="00:01:17:07" end="00:01:21:17">you can pick up the World Championship Pikachu Poké Plush</p>  
<p begin="00:01:21:19" end="00:01:23:02">at pokémoncenter.com.</p>
```

- **Step 1 - Sentence Forming**

And if that is not enough world championship action, you can pick up the World Championship Pikachu Poké Plush pokémoncenter.com.

# Strategy #1: Structure Forming

- **Subtitling DFXP Format**

```
<p begin="00:01:14:01" end="00:01:17:05">And if that is not enough world championship action,</p>  
<p begin="00:01:17:07" end="00:01:21:17">you can pick up the World Championship Pikachu Poké Plush</p>  
<p begin="00:01:21:19" end="00:01:23:02">at pokémoncenter.com.</p>
```

- **Step 1 - Sentence Forming**

And if that is not enough world championship action, you can pick up the World Championship Pikachu Poké Plush at pokémoncenter.com.

- **Step 2 - Translate**

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship Pikachu Poké Plush na pokémoncenter.com.

# Strategy #1: Structure Forming

- **Step 1 - Sentence Forming**

And if that is not enough world championship action, you can pick up the World Championship Pikachu Poké Plush at pokémoncenter.com.

- **Step 2 – Translate**

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship Pikachu Poké Plush na pokémoncenter.com.

- **Step 3 – Reformat & Re-insert**

```
<p begin="00:01:14:01" end="00:01:17:05">И если этого недостаточно для участия в чемпионате мира,</p>  
<p begin="00:01:17:07" end="00:01:21:17">вы можете забрать Pikachu Poké Plush Plush</p>  
<p begin="00:01:21:19" end="00:01:23:02">pokémoncenter.com.</p>
```

# Strategy #2: Zones & Walls

- **Moses decoder has the concept of**

- **Forced Translations** – parts of the input sentence can be wrapped in `<np>` tags and assigned a probability

```
<np translations="Translated Term" prob="1">...</np>
```

- **Zones**

- Encapsulates a series of tokens that cannot be individually re-ordered, but can be re-ordered as part of the parent sequence

```
You can pick up the World Championship <zone>Pikachu Poké  
Plush</zone>
```

- **Wall**

- A hard border that tokens cannot cross during re-ordering



# Strategy #3: Segmentation

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at [pokemoncenter.com](http://pokemoncenter.com).

# Strategy #3: Segmentation

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship

**<**

**Pikachu Poké Plush**

**>**

at pokémoncenter.com.

# Strategy #3: Segmentation

And if that is not enough world championship action, you can pick up the World Championship

`<strong>`

**Pikachu Poké Plush**

`</strong>`

at [pokemcenter.com](https://pokemcenter.com).

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship **Pikachu Poké Plush** na [pokemcenter.com](https://pokemcenter.com).

# Strategy #4: Pass Through

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at [pokemoncenter.com](http://pokemoncenter.com).

# Strategy #4: Pass Through

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship ❶ Pikachu Poké Plush ❷ at pokémoncenter.com.

Token	Assigned
❶	<strong>
❷	</strong>

# Strategy #4: Pass Through

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship ❶ Pikachu Poké Plush ❷ at pokémoncenter.com.

Token	Assigned
❶	<strong>
❷	</strong>

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship ❶Pikachu Poké Plush❷ na pokémoncenter.com.

# Strategy #4: Pass Through

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship ❶ **Pikachu Poké Plush** ❷ at pokémoncenter.com.

Token	Assigned
❶	<strong>
❷	</strong>

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship **Pikachu Poké Plush** na pokémoncenter.com.

# Strategy #4: Pass Through

Token	Assigned	Token	Assigned
①		⑥	
②		⑦	
③		⑧	
④		⑨	
⑤		⑩	



# Strategy #5: Optimisation

- **Detecting lead/trail tags**

```
<p><font="Calibri"><color="#1234"><size="10pts">And if that is not enough  
world championship action, you can pick up the World Championship  
<strong>Pikachu Poké Plush </strong>at pokémoncenter.com.  
</size></color></font></p>
```

# Strategy #5: Optimisation

- Detecting lead/trail tags

`<p><font="Calibri"><color="#1234"><size="10pts">And if that is not enough world championship action, you can pick up the World Championship <strong>Pikachu Poké Plush </strong>at pokémoncenter.com.</size></color></font></p>`

`<lead/>`And if that is not enough world championship action, you can pick up the World Championship **❶ Pikachu Poké Plush** **❷** at pokémoncenter.com.`<trail/>`

Token	Assigned
❶	<code>&lt;strong&gt;</code>
❷	<code>&lt;/strong&gt;</code>
Lead	<code>&lt;p&gt;&lt;font="Calibri"&gt;&lt;color="#1234"&gt;&lt;size="10pts"&gt;</code>
Trail	<code>&lt;/size&gt;&lt;/color&gt;&lt;/font&gt;&lt;/p&gt;</code>

# Which one to work with???

- **Strategy #1**
  - Structure Forming
- **Strategy #2**
  - Zones & Walls
- **Strategy #3**
  - Segmentation
- **Strategy #4**
  - Pass-Thru
- **Strategy #5**
  - Optimisation





**KantanMT.com**  
No Hardware. No Software. No Hassle MT.

## Challenges of Terminology

# Basic Levels of Terminology

- **Lexically Constrained Decoding for Sequence Generation**

- **Level I**

- Multi-word phrases – non-inflected
- Nouns, Proper Nouns, Brand or Feature/Service names
- Can be represented in simple list form

EN	FR	IT	DE
Stakataka	Ama-Ama	Stakataka	Muramura
Snorlax	Ronflex	Snorlax	Relaxo
Trainer	Dresseur	Allenatore	Trainer
Terrakion	Terrakium	Terrakion	Terrakium
Whimsicott	Farfaduvet	Whimsicott	Elfun
Gengar	Ectoplasma	Gengar	Gengar
Tapu Bulu	Tokotoro	Tapu Bulu	Kapu-Toro
Kommo-o	Ékaisser	Kommo-o	Grandiras

- **Level II**

- Inflected Multi-word phrases
- Cannot be represented in simple list form
- Inflected variations to indicate number, grammatical case, or gender

# What is a Glossary?

- **Terms in one or more languages**
  - With comments, and usage examples
- **Industry Glossary**
  - Includes terms that are standard for an industry or subject/domain
  - **Example:** Banking, Cardiology, Automotive, Accounting
- **Client Glossary**
  - Contains terms that are very specific to a company/client
  - **Example:** eBay, Adobe, BMW
- **Project Glossary**
  - Used to maintain consistency throughout translation project
  - **Example:** Microsoft Office, Pokemon



<https://iate.europa.eu/home>

**Terms:** 7.9 million  
**Languages:** 26

# Level #1: Segmentation

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship

<term>

**Pikachu Poké Plush**

</term>

at pokémoncenter.com.

# Level #1: Segmentation

And if that is not enough world championship action, you can pick up the World Championship

<term>

**Pikachu Poké Plush**

TermBase Lookup

</term>

at pokémoncenter.com.

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship **Pikachu Poké Plush** na pokémoncenter.com.



# Level #1: Pre-process & Pass-Thru

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at [pokémoncenter.com](http://pokémoncenter.com).

TermBase  
Lookup

And if that is not enough world championship action, you can pick up the World Championship `<np translations="Pikachu Poké Plush" prob="1"/>` at [pokémoncenter.com](http://pokémoncenter.com).

# Level #1: Pre-process & Pass-Thru

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

TermBase  
Lookup

And if that is not enough world championship action, you can pick up the World Championship `<np translations="Pikachu Poké Plush" prob="1"/>` at pokémoncenter.com.

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship **Pikachu Poké Plush** na pokémoncenter.com.



**KantanMT.com**  
No Hardware. No Software. No Hassle MT.

**Thank you...**