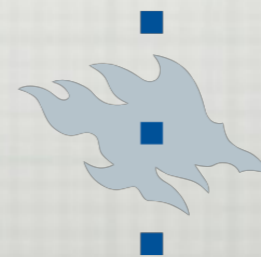


JOHDATUS TEKÖÄLYYN

TEEMU ROOS



HELSINGIN YLIOPISTO

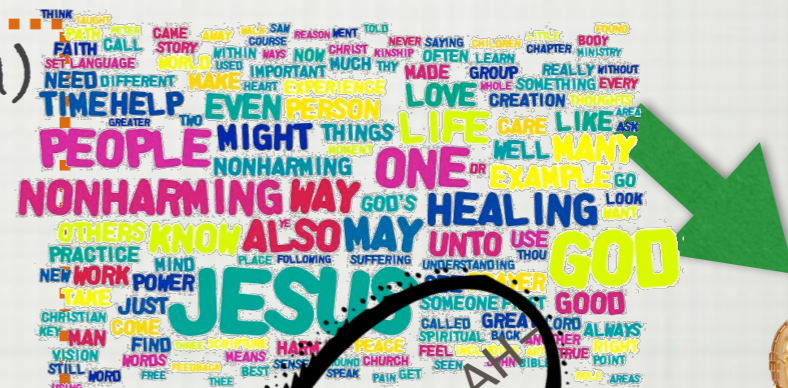
LUONNOLLISEN KIELEN KÄSITTELY (NATURAL LANGUAGE PROCESSING, NLP)

✳️ TEKOÄLYSOVELLUKSET, JOTKA LIITTYVÄT IHMISTEN KANSSA (TAI IHMISTEN VÄLISEEN) KOMMUNIKAATIOON, OVAT TEKEMISSÄ LUONNOLLISEN KIELEN KANSSA.

✳️ ESIM.:

- SUODATTAMINEN (ESIM. SPAM)
- LUOKITTELU
- TIEDONHAKU
- TIEDON KERÄÄMINEN
- KIELEN KÄÄNTÄMINEN
- TIIVISTÄMINEN
- KYSYMYKSIIN VASTAAMINEN
- DIALOGI

VAIKEUS



TEKSTI

- * TEKSTI POIKKEAA ESIM. DIGITAALISESTA SIGNAALIDATASTA
- * VAIKKA "1D" (ETENEE YHTEEN SUUNTAAN), RAKENNE HIERARKKINEN (TEKSTI, KAPPALE, LAUSE, SANA)
- * KIELIOPPI RAJOITTA MUOTOA (MUTTA LUONNOLLINEN KIELI SILTI MONISELITTEISTÄ)
- * LÄHEMPÄNÄ SEMANTIikkaa KUIN DIGITAALISET SIGNAALIT:

"TUOLI" ⇔ TUOLIN IDEA ⇔ ?



KIELET JA KIELIOPIT

* TIETOJENKÄSITTELYTIETEEN TEORIASSA **FORMAALI KIELI** TARKOITTA A MERKKIJONOJEN JOUKKOA

* ESIM. SALLITUT MATEMAATTISET KAAVAT:

$(1+2) \times 6 - 3$

KUULUU KIELEEN

$1 + (+ \times 5 ($

EI KUULU KIELEEN

* **YHTEYDETÖN KIELI** (CONTEXT FREE LANGUAGE) = KIELI, JOKA VOIDAAN TUOTTAA YHTEYDETTÖMÄLLÄ **KIELIOPILLA**:

JOUKKO SÄÄNTÖJÄ

$V \rightarrow w$

VÄLIKE

yhteydetön = ei riipu ympäröivästä kontekstistä

JONO VÄLIKKEITÄ TAI PÄÄTESYMBOLIJA

KIELET JA KIELIOPIT

- * ESIMERKKI YHTEYDETTÖMÄSTÄ KIELIOPISTA:

$S \rightarrow SS$

$S \rightarrow$

$S \rightarrow (S)$

$S \rightarrow [S]$

S : ALOITUSSYMBOLI

$() []$: PÄÄTESYMBOLIT

- * $S \Rightarrow SS \Rightarrow SSS \Rightarrow SSSS \Rightarrow (S)SSS \Rightarrow (S)(S)SS \Rightarrow (S)(S)(S)S$
 $\Rightarrow (S)(S)(S)(S) \Rightarrow ()(S)(S)(S) \Rightarrow ()()(S)(S) \Rightarrow ()()()(S) \Rightarrow ()()()()$

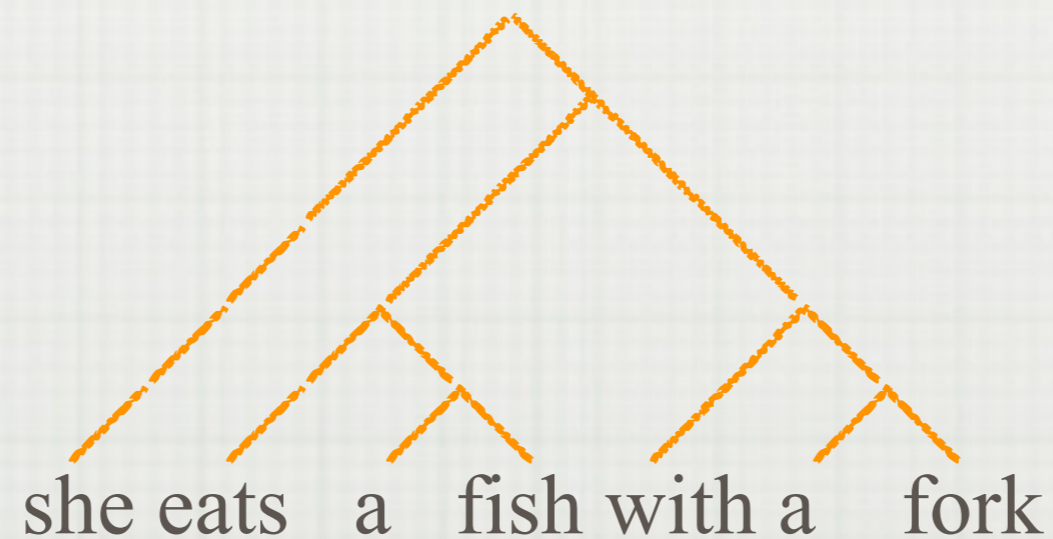
- * KUULUU KIELEEN: $"()"$, $"()()()"$, $"((()))"$, $"([()])"$, ...

- * EI KUULU KIELEEN: $"((", "((", "(]", "([()])"$, ...

KIELET JA KIELIOPIT

* KIELEN TUNNISTUS: JÄSENENNYYS

* JÄSENENNYSPUU

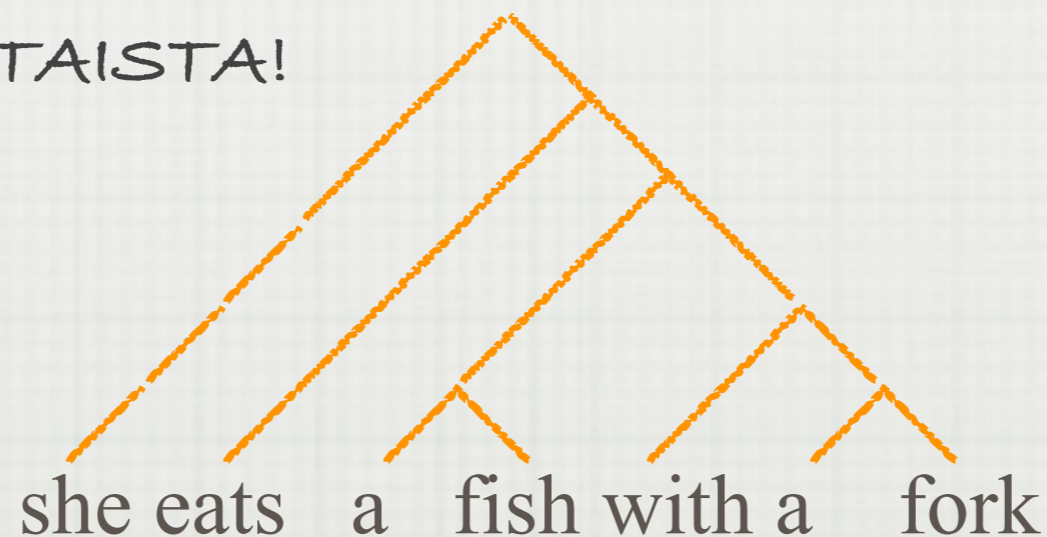


KIELET JA KIELIOPIT

* KIELEN TUNNISTUS: JÄSENENYS

* JÄSENENYSPUU

* MONITULKINTAISTA!



* SÄÄNTÖIHIN VOIDAAN LIITTÄÄ TODENNÄKÖISYYDET

* LEKSIKAALISTETTU JÄSENENYS: SEMANTIikka MUKAAN

CYK-ALGORITMI

- * COCKE-YOUNGER-KASAMI (1965-1970)
- * PERUSVERSIO TUNNISTAA YHTEYDETTÖMÄN KIELEN
- * PERUSTUU DYNAAMISEEN OHJELMOINTIIN (HAJOTAJA HALLITSE)
- * VOIDAAN HELPOSTI LAAJENTAA TUOTTAMAAN TODENNÄKÖISINJÄSENNYSPUUN

CYK-ALGORITMI

$S \rightarrow NP VP$

$VP \rightarrow VP PP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$PP \rightarrow P NP$

$NP \rightarrow Det N$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$N \rightarrow she$

$N \rightarrow fish$

$N \rightarrow fork$

$V \rightarrow eats$

$V \rightarrow fish$

$P \rightarrow with$

$Det \rightarrow a$

1. ALOITA TYHJÄSTÄ TAULUKOSTA

2. KÄSITTELE RIVIT ALHAALTA YLÖS:

3. KUNNES EI VOI SOVELTAA SÄÄNTÖÄ:

4. SOVELLA SÄÄNTÖJÄ ALEMPIEN RIVIEN SYMBOLEIHIN

$T(i,j) \rightarrow T(i,j)$

$T(i,j) \rightarrow T(i,k) T(k+1,j)$

(1,7)						
(1,6) (alku, loppu)						
(1,5)	(2,6)	(3,7)				
(1,4)	(2,5)	(3,6)	(4,7)			
(1,3)	(2,4)	(3,5)	(4,6)	(5,7)		
(1,2)	(2,3)	(3,4)	(4,5)	(5,6)	(6,7)	
(1,1)	(2,2)	(3,3)	(4,4)	(5,5)	(6,6)	(7,7)
she	eats	a	fish	with	a	fork

CYK-ALGORITMI

$S \rightarrow NP VP$

$VP \rightarrow VP PP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$PP \rightarrow P NP$

$NP \rightarrow Det N$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$N \rightarrow she$

$N \rightarrow fish$

$N \rightarrow fork$

$V \rightarrow eats$

$V \rightarrow fish$

$P \rightarrow with$

$Det \rightarrow a$

1. ALOITA TYHJÄSTÄ TAULUKOSTA

2. KÄSITTELE RIVIT ALHAALTA YLÖS:

3. KUNNES EI VOI SOVELTAA SÄÄNTÖÄ:

4. SOVELLA SÄÄNTÖJÄ ALEMPIEN RIVIEN SYMBOLEIHIN

$T(i,j) \rightarrow T(i,j)$

$T(i,j) \rightarrow T(i,k) T(k+1,j)$

N,NP	V,VP	Det	N,NP,V,VP	P	Det	N,NP
she	eats	a	fish	with	a	fork

CYK-ALGORITMI

$S \rightarrow NP VP$

$VP \rightarrow VP PP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$PP \rightarrow P NP$

$NP \rightarrow Det N$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$N \rightarrow she$

$N \rightarrow fish$

$N \rightarrow fork$

$V \rightarrow eats$

$V \rightarrow fish$

$P \rightarrow with$

$Det \rightarrow a$

1. ALOITA TYHJÄSTÄ TAULUKOSTA

2. KÄSITTELE RIVIT ALHAALTA YLÖS:

3. KUNNES EI VOI SOVELTAA SÄÄNTÖÄ:

4. SOVELLA SÄÄNTÖJÄ ALEMPIEN RIVIEN SYMBOLEIHIN

$T(i,j) \rightarrow T(i, j)$

$T(i,j) \rightarrow T(i, k) T(k+1, j)$

$T(1,2) \rightarrow T(1,1) T(2,2)$

S						
N,NP	V,VP	Det	N,NP,V,VP	P	Det	N,NP
she	eats	a	fish	with	a	fork

CYK-ALGORITMI

$S \rightarrow NP VP$

$VP \rightarrow VP PP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$PP \rightarrow P NP$

$NP \rightarrow Det N$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$N \rightarrow she$

$N \rightarrow fish$

$N \rightarrow fork$

$V \rightarrow eats$

$V \rightarrow fish$

$P \rightarrow with$

$Det \rightarrow a$

1. ALOITA TYHJÄSTÄ TAULUKOSTA

2. KÄSITTELE RIVIT ALHAALTA YLÖS:

3. KUNNES EI VOI SOVELTAA SÄÄNTÖÄ:

4. SOVELLA SÄÄNTÖJÄ ALEMPIEN RIVIEN SYMBOLEIHIN

$T(i,j) \rightarrow T(i,j)$

$T(i,j) \rightarrow T(i,k) T(k+1,j)$

			NP (3,7)			
S				NP,VP		
	VP				PP (5,7)	
S			NP (3,4)			NP
N,NP	V,VP	Det	N,NP,V,VP	P	Det	N,NP
she	eats	a	fish	with	a	fork

$T(3,7) \rightarrow T(3,4) T(5,7)$

CYK-ALGORITMI

$S \rightarrow NP VP$

$VP \rightarrow VP PP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$PP \rightarrow P NP$

$NP \rightarrow Det N$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$N \rightarrow she$

$N \rightarrow fish$

$N \rightarrow fork$

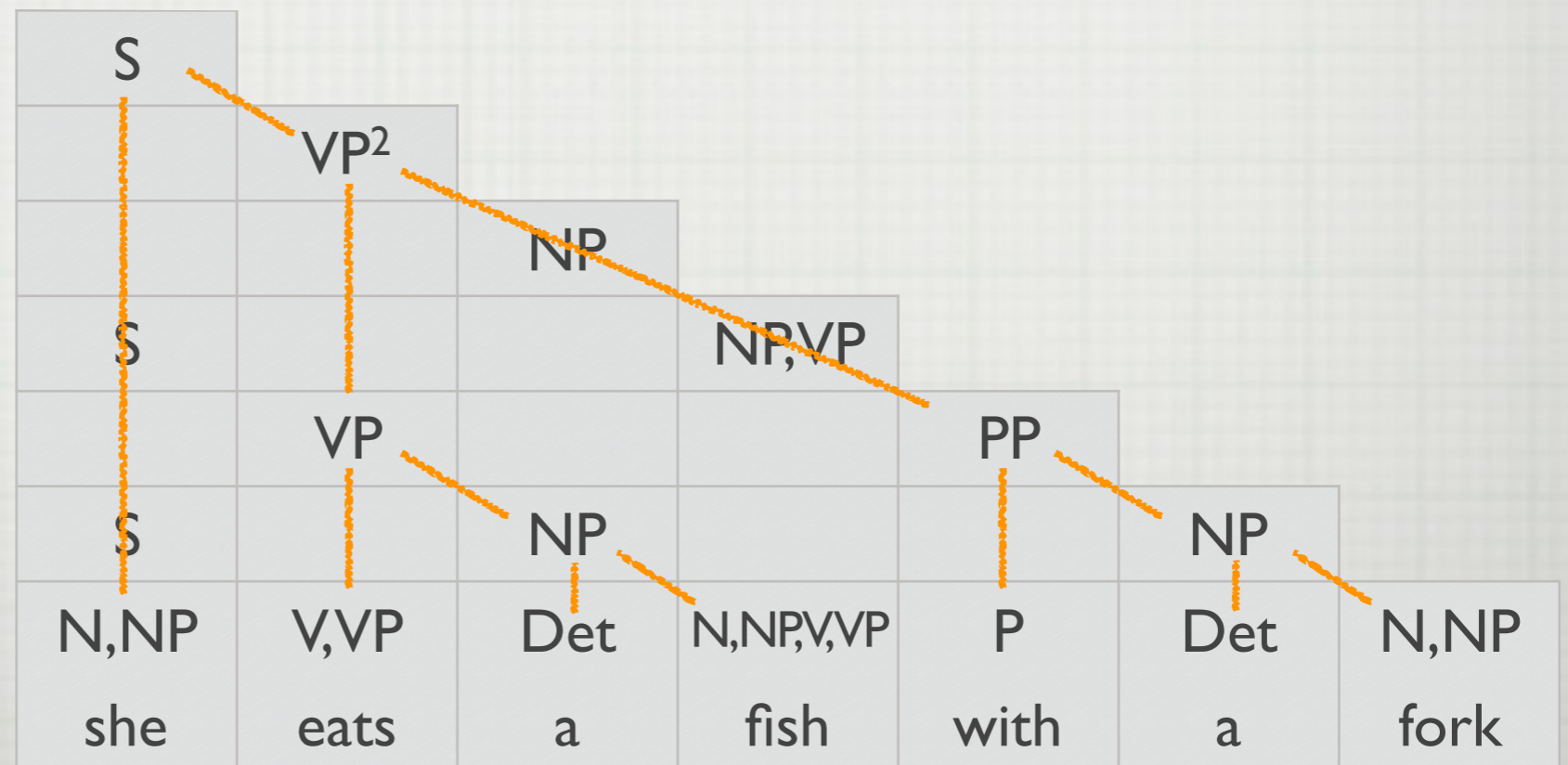
$V \rightarrow eats$

$V \rightarrow fish$

$P \rightarrow with$

$Det \rightarrow a$

JÄSENNYSPUUT RAKENNETTAVISSA
TAULUKOSTA



CYK-ALGORITMI

$S \rightarrow NP VP$

$VP \rightarrow VP PP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$PP \rightarrow P NP$

$NP \rightarrow Det N$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$N \rightarrow she$

$N \rightarrow fish$

$N \rightarrow fork$

$V \rightarrow eats$

$V \rightarrow fish$

$P \rightarrow with$

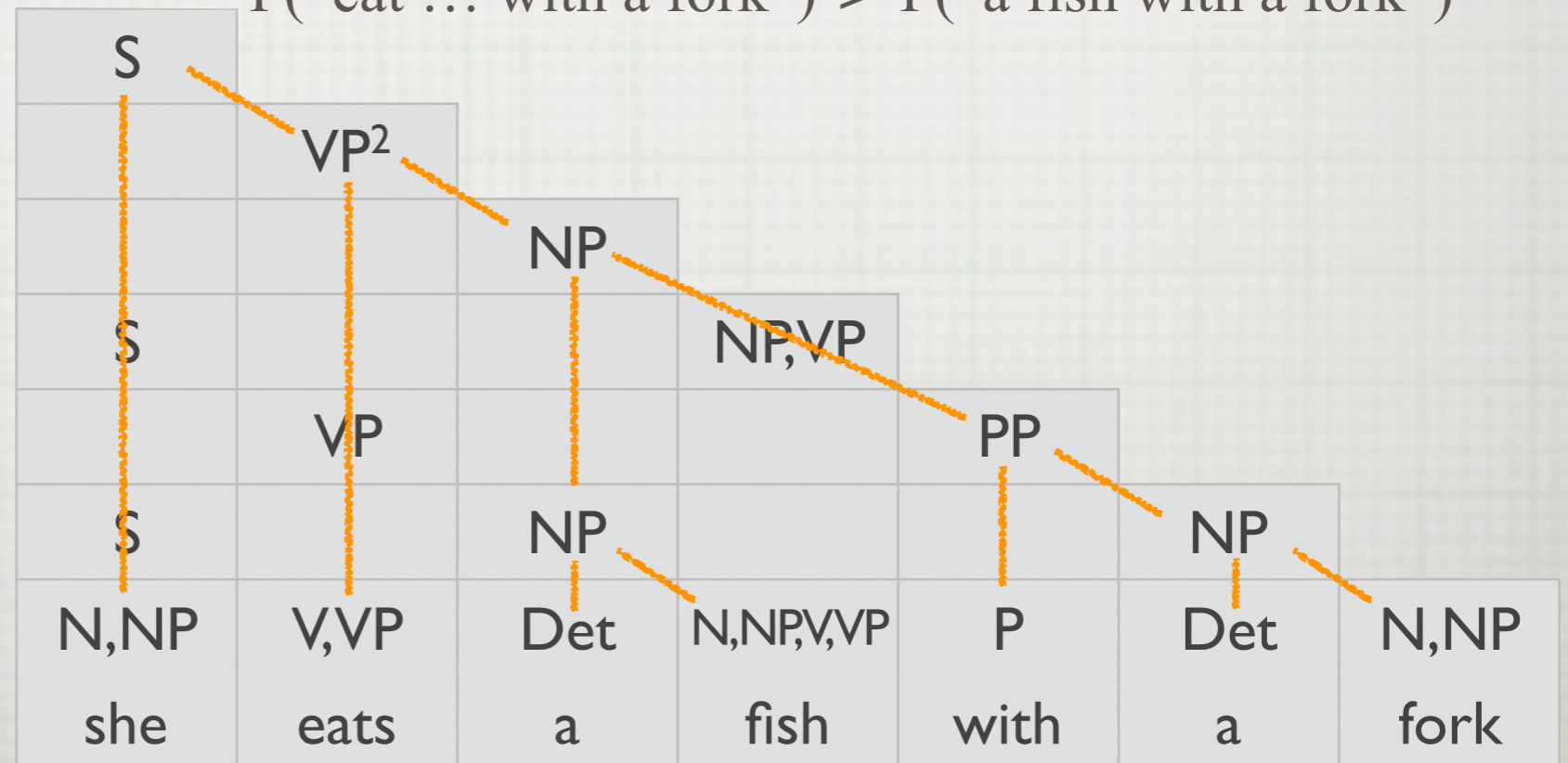
$Det \rightarrow a$

JÄSENNYSPUUT RAKENNETTAVISSA
TAULUKOSTA

LEKSIKALISOITU JÄSENNYYS:

TODENNÄKÖISIN JÄSENNYYS RIIPPUU
SEMANTIIKASTA

$P(\text{"eat ... with a fork"}) > P(\text{"a fish with a fork"})$



SOVELLUKSIA

* JÄSENYYKSEN AVULLA VOIDAAN:

○ TUNNISTAA RELAATIOITA, ESIM.

"JEAN SIBELIUS SÄVELSI VALSE TRISTEN."

⇒ Säveltäjä(Jean Sibelius, *Valse Triste*)

"GOOGLE ACQUIRES ARTIFICIAL INTELLIGENCE STARTUP

DEEPMIND FOR MORE THAN \$500M"

⇒ Omistaa(Google, Deep Mind)

○ AVITTAAM KONEKÄÄNNÖSTÄ

○ VOITTAAM JEOPARDY!

SOVELLUKSIA

* JÄSENNYKSEN AVULLA VOIDAAN:

- LOUHIA NUMEERISTA DATAA SUURESTA MÄÄRÄSTÄ
KUTISIA: PULS

SA treats 68 cholera patients on Zimbabwe border

South Africa has treated **68 cholera patients** since the weekend in a town by the border with **Zimbabwe**, where the **disease has killed dozens** of people **in recent weeks**, a health official said today. "**Since Saturday**, we have received and treated a total of **68 cholera patients from Zimbabwe**," said Phuti Selobi, spokesman for the health department in the town of Musina said. "Sixty-six of them are Zimbabweans while two others are South Africans engaged in cross-border business," Selobi told AFP. "Only 14 of them are still in the hospital," he added, noting that no one has died of cholera in South Africa.

Understanding

Each detected case is transformed into **structured information**:

- **what** disease
- **where** — country, location,
- **when** — date
- **who** — victims, number, ...

and added to a database of events



Disease	Cholera
Country	Zimbabwe
Location	Musina
Time	2008-11-20
Victim	People
Number	68

Learning

PULS gives authorized users the option to provide feedback to the system, e.g., by correcting errors in the automatic analysis. The system adapts to improve the analysis on subsequent reports.