

# Automatický překlad přirozených jazyků

# Kageru-to mugade hala jo deka medsene seno gejay!

kageru = sejít se, shromáždit se, oženit se

mugade = hlava

hala = skupina

jo = jeden, jedna

deka = pět

medse = hodina

seno = velký

jay = muž, vládce

Překlad:

*Sejít se hlava skupina jeden pět ?hodina?  
velký (?muž, vládce?).*

# Co nám chybí?

## Tvarosloví (morfologie)

*<Sloveso>* + to = Budoucí čas

*Přípona* -ne = 6.pád (časová předl. v)

*Předpona* ge- = 9.pád (místní určení)

*Sejde se hlava skupina jeden pět v  
hodině velký u muže.*

# "Pravidla domorodého pravopisu"

- v oznamovacích větách domorodci používají pevný pořádek slov, kde se za slovesem ostatní větné členy seřadí podle své důležitosti, počínaje podmětem.
- systém číslovek se podobá římským číslům - *jeden pět* znamená *čtyři*.
- přídavná jména, zájmena a číslovky vždy rozvíjejí bezprostředně následující jména podstatná

*Sejde se hlava skupina ve čtyři hodiny u velkého muže*

## Ustálená spojení

**mugade hala** se má překládat buď jako *skupina moudrých* nebo ještě lépe jako *rada starších*

**seno jay** není *velký muž*, ale *náčelník*

*Rada starších se sejde ve čtyři hodiny u náčelníka*

# Problémy překladu – rozdíly ve významu

anglicky			japonsky		
cook	bake		YAKU		↑ Pouze žár
	toast				
	roast				
	broil	barbecue			
		grill	ABURU		↓
		stir-fry sauté			
	fry	deep-fry French-fry	AGERU FRY NI SURU TEMPURA NI SURU		↑ Vaření na oleji
		braise			
	boil	simmer	NIRU TAKU		↑ Vaření ve vodě
		stew poach			
		boil	YUDERU		
		steam	MUSU, FU		

## Víceznačnost slov

Remove the spark plugs	Zündkerzen herausdrehen
Remove the plug leads	Zündkabel abziehen
Remove the dipstick	Öhlmeßstab herausziehen
Remove the filter cap	Verschlußkappe aufdrehen
Remove the distributor cap	Verteildeckel abnehmen
Remove the rotor arm	Verteilerläufer abziehen
Remove nipple	Schmiernippel herausdrehen
Remove the two bolts	Beide Schrauben lösen

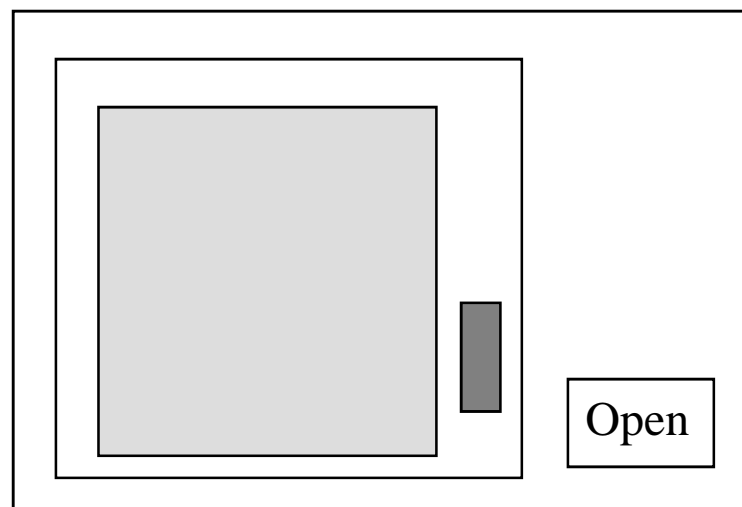
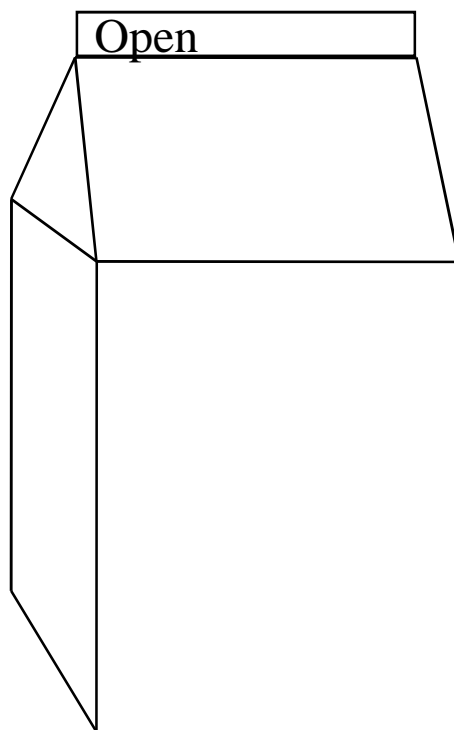
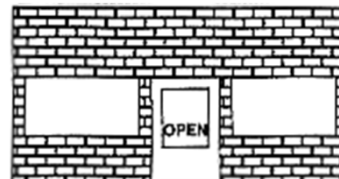
## Složené výrazy

... remove defective *left wing tip tank filler gap gasket*...

... airport long term car park courtesy vehicle pickup point ...

Mädchenhandelschule

# Závislost na situaci



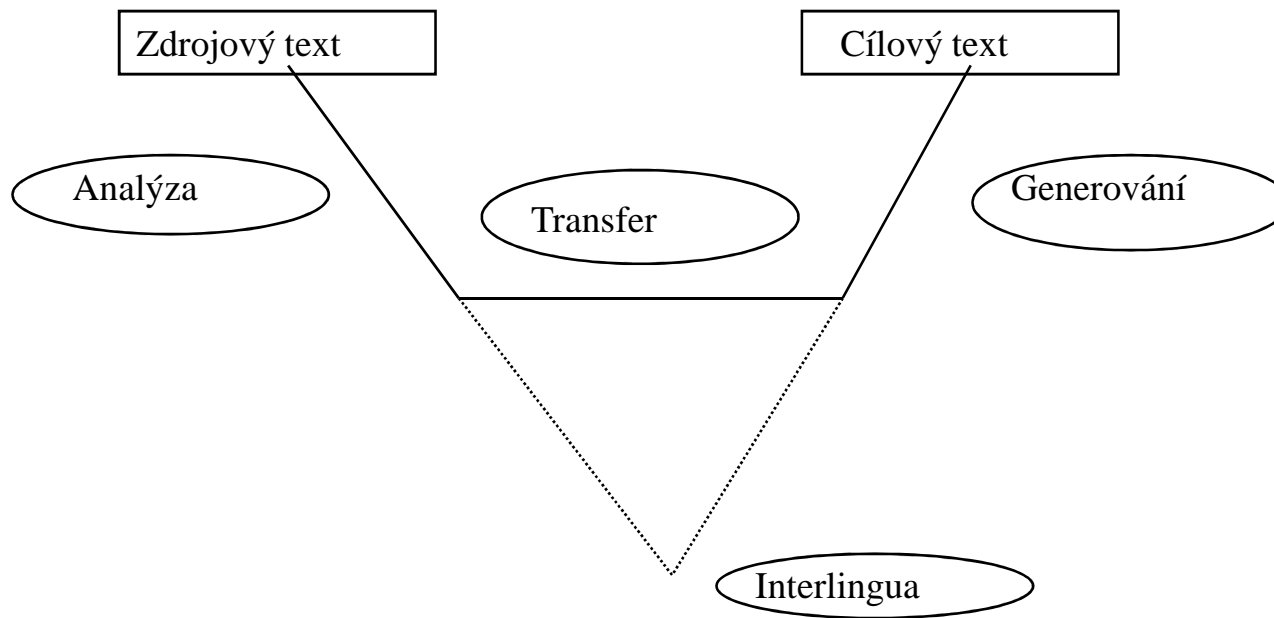


# Prehistorie

1933 patenty pro překladové stroje:

George Artsruni

P.P Smirnov-Trojanskij



# První pokusy

1946 A.D.Booth - idea automatického dvojjazyčného slovníku, text zpracováván slovo od slova

1948 R.M.Richens - ve slovníku nejsou zachycena celá slova, ale předpony, kmeny a přípony zvlášť

1950 E.Reifler - zavádí pre- a post- editing

1952 První konference o strojovém překladu na MIT

L.E.Dostert - pivotní jazyk pro překlad více jazyků (přirozený jazyk, sloužící jako jakýsi „most“ mezi zdrojovým a cílovým jazykem)

7.1.1954 Georgetownský experiment (do r. 1956)

45 vět s 250 slovy, 6 syntaktických „zákonů“, jednoduché oznamovací věty, bez negací, slovesa ve 3.osobě, málo předložek

# Bouřlivý rozvoj

1955 Anglo-ruský překlad v Moskvě

1956 První mezinárodní konference

12 vědeckých skupin na amerických univerzitách

1957 N.Chomsky - Standard theory

1960 Y.Bar Hillel: „Vysoce kvalitní plně automatický překlad nemůže být nikdy dosažen.“

„The box was in the pen.“

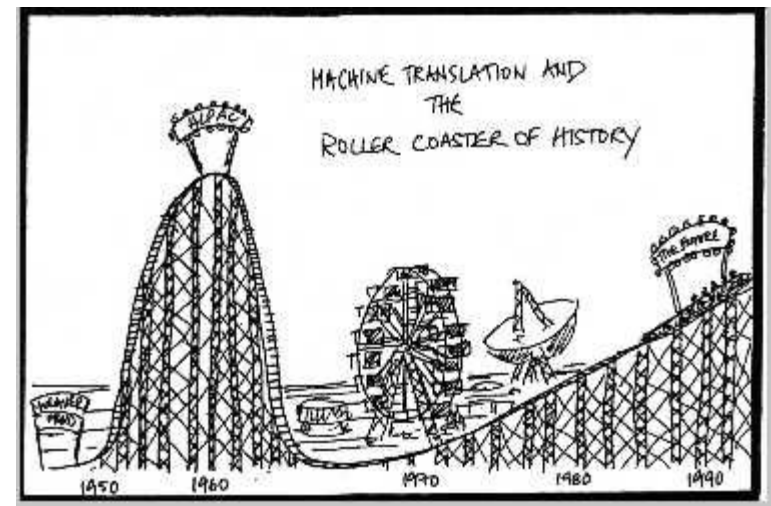
Příklon k teorii, práce se syntaxí jazyka, nová generace systémů

# Vystřízlivění

1966 Zpráva **ALPAC** (American Language Processing Advisory Committee)

- Konstatuje nutnost investic do dlouhodobého teoretického lingvistického výzkumu.
- Faktický důsledek: konec podpory

Zpráva nebyla negativní, negativní byly její důsledky



Práce mimo USA pokračovaly (Francie, SSSR, Kanada)

# Období po ALPACu

„Doba temna“

Existují výjimky:

- SYSTRAN (konec 60.let)
- skupina GETA v Grenoblu (vznikla 1961)
- SUSY (Saarbruecken)
- LOGOS (Texas)
- TAUM (Montreal)
- ETAP (Moskva)

# První úspěšný komerční systém:

## TAUM - METEO (1976)

- překlad meteorologických zpráv A->F
- dobře definovaná a výrazně syntakticky i sémanticky omezená podmnožina jazyka
- vhodné implementační prostředky
- systém sám rozezná text, který mu dělá potíže, a předá jej lidskému překladateli
- v 90.letech překládali 45 000 slov denně

# Další významné systémy I.

## SYSTRAN

- Překlad dokumentů EU
- Přímý překlad (každý pár řešen zvlášť) mezi cca 20 páry, ovšem uspokojivá kvalita pouze u nejstarších párů (A-F-N)
- Data oddělena od programu
- Problémy řešeny ad hoc

## EUROTRA

- Oficiální projekt EU v 80.letech
- Megalomanie: 72 jazykových párů
- Nezvládnutá modularita
- Do jisté míry podobný negativní efekt jako zpráva ALPAC

# Další významné systémy II.

## VERBMOBIL

- Německý nástupce EUROTRy
- Překlad mluvené řeči
- Tématické omezení rozhovoru - plánování příští schůzky dvou obchodníků
- Více než 30 výzkumných týmů
- Předváděn na Světové výstavě v Hannoveru, od té doby se o něm příliš nepíše.



# Současné trendy

## Statistické metody

- využívají paralelní označované korpusy
- rychle se zlepšují, pro měření kvality však používají technické metody založené na referenčních překladech – BLEU score

## Nástroje podporující překlad

Systémy využívající dříve přeložených textů - princip tzv. překladové paměti

IBM Translation Manager, Dejà Vu

**TRADOS Translator's Workbench**

SDLX

Vhodné zejména pro lokalizace dokumentace k systémům, které vycházejí ve stále aktualizovaných verzích

# Exportovaná překladová paměť

<RTF Preamble>

<FontTable>

{\fonttbl

{\f1 \fmodern\fprq1 \fcharset0 Courier New;}

{\f2 \fswiss\fprq2 \fcharset0 Arial;}

{\f3 \froman\fprq2 \fcharset2 Symbol;}

{\f4 \froman\fprq2 {\\*\falt Times}\fcharset0 Times New Roman;}

{\f5 \froman\fprq2 {\\*\falt Times}\fcharset0 Times New Roman CE;}

...

<TrU>

<CrD>04021999

<CrU>SAP1

<Att L=Subunit>DOC

<Txt L=Text Field>ASAP\_2/99

<Seg L=CS\_01>Zahrňte informace o tom, jaká bude spolupráce technického a aplikačních týmů).

<Seg L=PL\_01>Włączcie informacje o tym, jaka będzie współpraca technicznego a aplikacyjnych teamów).

</TrU>

# České systémy

---

První věta přeložena v r.1957 na počítači SAPO

„The consonants have not by far been investigated to the same extent as the vowels.“ -

Výsledek:

„Souhlásky zdaleka nebyly prozkoumány do stejné míry jako samohlásky.“

Existující systémy:

- **APAČ** (80. léta)
- **RUSLAN** (1985-1990)
- **ČESÍLKO** (1998-?)

# System APAČ

---

- Překlad z angličtiny do češtiny
- Autorem Dr. Zdeněk Kirschner
- Vytvořen ve shodném formalismu jako METEO
- Překlad textů z oblasti vodních pump
- Transdukční slovník:
  - ation => -ace
  - industrialization => industrializace
  - ic => -ický
  - static => statický
- Slovník s cca 1500 výrazy

# System RUSLAN

---

- Překlad manuálů k operačním systémům sálových počítačů
- Implementace na PC AT 286
- Překlad 1 věty cca 4 minuty
- Slovník o velikosti cca 8500 slov
- Transdukční slovník
- Minimální transfer (stále ale rostl)
- Gramatika zapsána pomocí Q-systémů (TAUM)
- Záchranná pravidla pro případ problémů při analýze
- Výzkum zastaven v roce 1990 těsně před provozními zkouškami

# Formát dat

---

## Vstup:

-1- \$sn(5) + \$gs(1) + < + Urc3ity2 + investic3ni2 + majetek + je +  
vz3dy + jednoznac3ne3 + pr3ir3azen + te3mto + organizac3ni2m +  
jednotka2m + . + >-2-

## Slovník:

PR3IR3ADI==R(5,PRM,?(N(N),A(A),/,N(N),I(I),/,>,</,D(D)),15,PRISVOJIT6).

## Výsledek:

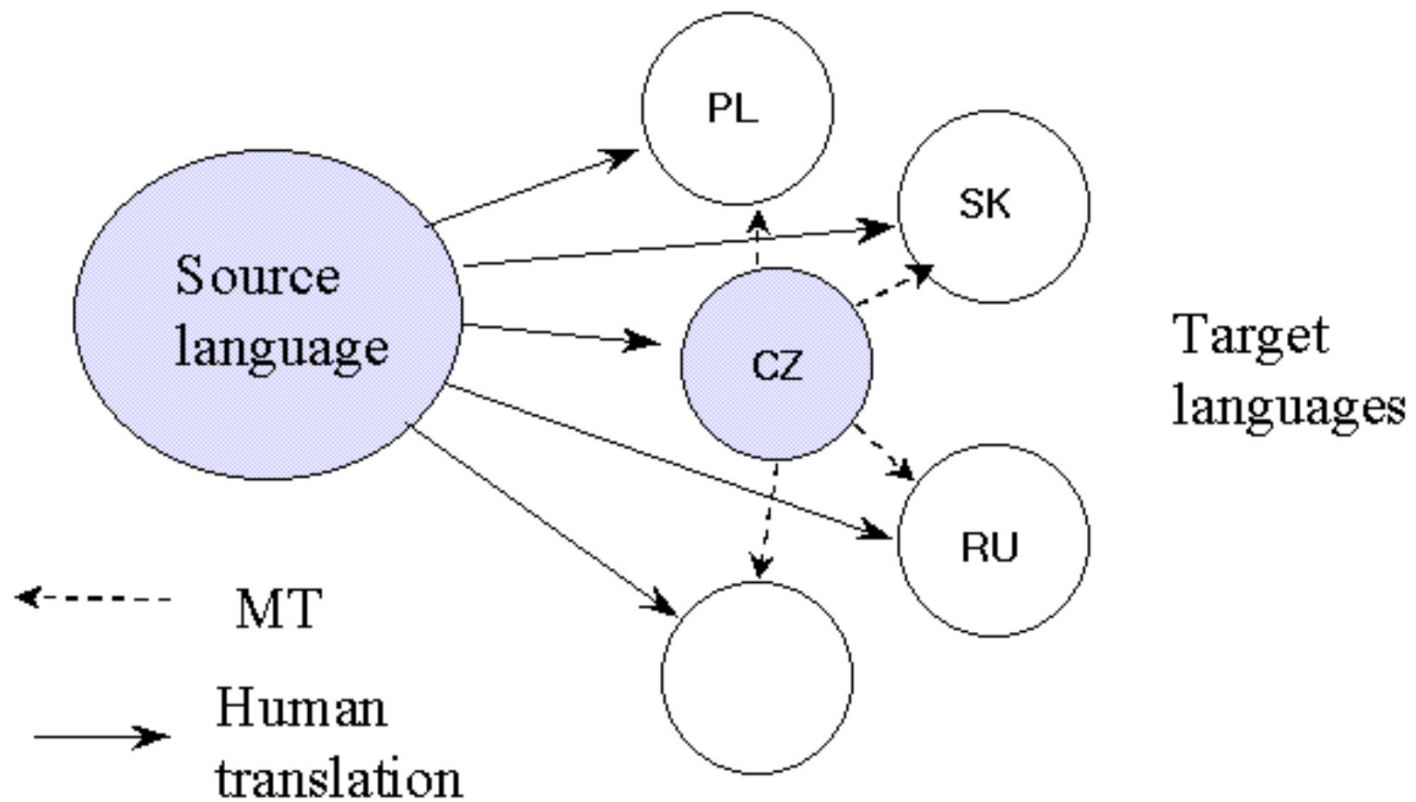
OPREDELENNOE INVESTICIONNOE IMUW"ESTVO VSEGDA  
ODNOZNAC"NO PRISVOJENO TOT Z"E ORGANIZACIONNYM  
EDINICAM !SEGMENTAM .

# System Česílko

---

- Lokalizace velkých softwarových systémů
- Snaha minimalizovat lidský podíl na překladu
- Náhrada lokalizace z jazyka typově odlišného překladem z jazyka blízkého
- Typ překladu:
  - FAHQ (Plně automatický, vysoce kvalitní)
  - Velmi blízké jazyky
  - Plné morfologické slovníky
  - Statistická analýza češtiny
- Ověření možnosti použití shodné technologie pro jiné páry (skupiny) jazyků

## Využití příbuznosti





# Příklad překladu / Rozdíly

---

- Z češtiny do slovenštiny:
  - Při zakládání třídy výkazů se *třídě* nejprve přidělí označení a *přiřadí* se skupině uživatelů.  
=====>
  - Pri zakladaní triedy výkazov sa *triede* najprv prideli označenie a *priradí* sa skupine užívateľov.
- Shodná syntaxe
- Většinou shodné pořadí slov ve větě
- Odlišné slovníky (jistá pravidelnost)
- Naprosto odlišné tvarosloví

# Výsledky překladu CZ-SK

## Originál:

Výklad metodou postupných kroků o stávající situaci podnikového procesu, která existuje ve vašem *podnikání*.

Představuje popis formou scénáře, v němž normálně popíšete skutečnou situaci, ve které se nachází vaše podnikání. Nemusíte se zabývat podrobnými vstupními nebo výstupními požadavky, různými podmínkami nebo *očekávanými* hodnotami, neboť to je provedeno v kapitolách o konfiguraci a *testování*. Popis případu by měl být konzistentní s tokem podnikového procesu a měl by plně podporovat tok podnikového procesu, který je pro tento případ v *Základním seznamu podnikových procesů*. Jak se použijí tyto informace: jsou extrapolovány do specifikací konfigurace, které se použijí při konfiguraci vašeho systému R/3.

## Překlad:

Výklad metódou postupných krokov o súčasnej situácii podnikového procesu, ktorá existuje vo vašom *podnikaníe*.

Predstavuje popis formou scenára, v ktorom normálne popíšete skutočnú situáciu, vo ktorej sa nachádza vaše podnikanie. Nemusíte sa zaoberať podrobnými vstupnými alebo výstupnými požiadavkami, rôznymi podmienkami alebo *očakávať* hodnotami, pretože to je vykonané v kapitolách o konfigurácii a *testovanie*. Popis prípadu by mal byť konzistentný s tokom podnikového procesu a mal by plne podporovať tok podnikového procesu, ktorý je pre tento prípad v *Business Process Master listov*. Ako sa použijú tieto informácie: sú extrapolované do špecifikácií konfigurácie, ktoré sa použijú pri konfigurácii vášho systému R/3.

# Výsledky překladu CZ-PL

## Originál:

Výklad metodou postupných kroků o stávající situaci podnikového procesu, která existuje ve vašem *podnikání*.

Představuje popis formou scénáře, v němž normálně popíšete skutečnou situaci, ve které se nachází vaše podnikání. Nemusíte se zabývat podrobnými vstupními nebo výstupními požadavky, různými podmínkami nebo *očekávanými* hodnotami, neboť to je provedeno v kapitolách o konfiguraci a *testování*. Popis případu by měl být konzistentní s tokem podnikového procesu a měl by plně podporovat tok podnikového procesu, který je pro *tento* případ v *Základním seznamu podnikových procesů*. Jak se použijí tyto informace: jsou extrapolovány do specifikací konfigurace, které se použijí při konfiguraci vašeho systému R/3.

## Překlad:

Wystawa metodą postępowanych kroków o obecnej sytuacji zakładowego procesu, która istnieje w waszym *przedsiębiorstwo*.

Przedstawia opis formą scenariusza, w którym normalnie opiszeć rzeczywistą sytuację, w której się znajduje wasze przedsiębiorstwo. Nie Musicie się zajmować szczegółowymi wejściowymi czy zadanymi wyjściowymi, różnymi warunkami czy wartościami *oczekiwaniemi*, albowiem to jest wykonano w kapitułach o konfiguracji a *testowanie*. Opis wypadku miałby być niespójny z płynięcie zakładowego procesu a miałby do pełna wspierać płynięcie zakładowego procesu, który jest dla *ten* wypadek w *Liście głównym procesów biznesowych*. Jak się użyją te informacje: są ekstrapolacja do wyszczególnień konfiguracji, które się użyją przy konfiguracji waszego systemu R/3.

# České komerční systémy

---

Marsh & McLennan Cos. said it agreed to acquire the rest of Gradmann & Holler, a leading West German insurance brokerage firm in which it has held a 15% stake for 15 years. The transaction, for cash and stock, would represent the biggest European takeover since 1980 for New York-based Marsh & McLennan, the world's largest insurance broker. It's also the first major sign of the long-awaited consolidation in the European insurance industry as the European Community Commission moves toward a single market by 1992

## PC Translator 2003

Mokřina & McLennan kosinusy. řekly, že to souhlasilo se získat zbývající část Gradmann & křičí, vůdčí západní německý pojišťovací makléřská firma v kterém to držela 15% sázka pro 15 let. Transakce, za hotové a akcie, reprezentovala největší evropské převzetí od roku 1980 pro New York-založená mokřina & McLennan, světově největší pojišťovací agent. To je také první významnější znamení dlouho očekávaného upevnění v evropském pojišťovnictví jak Evropské společenství pověřovací pohyby směrem k jednotnému trhu do 1992.

# Google Translate

---

Marsh & McLennan Cos. said it agreed to acquire the rest of Gradmann & Holler, a leading West German insurance brokerage firm in which it has held a 15% stake for 15 years. The transaction, for cash and stock, would represent the biggest European takeover since 1980 for New York-based Marsh & McLennan, the world's largest insurance broker. It's also the first major sign of the long-awaited consolidation in the European insurance industry as the European Community Commission moves toward a single market by 1992

Marsh & McLennan dohodnuta Cos. jí řekl, že k získání zbytku Gradmann & křičet, přední pojišťovací makléřství západoněmecké firmy, v nichž má v držení 15% akcií za 15 let. Tato transakce, za výběr hotovosti a akciích, by představovalo největší evropský převzetí od roku 1980 v New Yorku založena-Marsh & McLennan, který je největším na světě pojišťovací makléř. Je také prvním hlavním znakem je dlouho očekávaná-konsolidace v evropském pojišťovnictví, jako je Evropská společenství Komise podniká kroky směrem k vytvoření jednotného trhu do roku 1992